DOCUMENT RESUME

ED 278 692 TM 870 109

AUTHOR Cason, Gerald J.; And Others
TITLE A Flexible System for Processing Clinical Performance Ratings: Illustrative Applications in a Residency and Four Clerkships.
PUB DATE Oct 86
NOTE 42p.; Paper presented at the Annual Meeting of the Association of American Medical Colleges (New Orleans, LA, October 25-30, 1986). Figures contain small print.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Clinical Teaching (Health Professions); *Computer Oriented Programs; Data Collection; Educational Assessment; *Evaluation Methods; Graduate Medical Education; Measures (Individuals); Medical Education; *Medical Evaluation; *Rating Scales; Record Keeping; Records (Forms); *Reports; Scoring; *Student Evaluation
IDENTIFIERS FORTRAN Program Language; University of Arkansas for Medical Sciences

ABSTRACT

The Performance Rating portion of the University of Arkansas for Medical Sciences (UAMS) Objective Test Scoring and Performance Rating (OTS-PR) system is designed to help the clinical teacher evaluate students' clinical performance. The system collects, processes, and reports ratings of learners' performances in clinical settings. The system's 160 modular FORTRAN programs are presently run on a mainframe computer, but since the user interacts with an Assessment Scoring Coordinator not the computer, the user (e.g., course director) need have no technical knowledge of computers, printers, or scanners. The system maintains records for up to 30 separate assessment occasions (tests or ratings) per course. Each user may specify up to 40 rating criteria per inventory and 30 inventories with different criteria. The rating criteria used in one course need have no relationship to those used in another. The paper first provides a detailed, generic description of the system followed by descriptions of its application in a Pediatrics Residency and Junior Year medical school clerkships (i.e., clinically oriented courses) in Pediatrics, Sub-specialties, Surgery, and Obstetrics/Gynecology. Each application section contains an example inventory, reports containing actual data, and the user's comments on the usefulness of the system. Experience indicates that the system has greatly improved the quality, quantity, completeness, and timeliness of evaluation information available to educational administrators, faculty, and students alike. The system would likely prove useful in any training or practice setting, health-care or otherwise, where repeated ratings of human performance by multiple raters is needed to insure proper quality control. (Author/LMO)

ED278692

A Flexible System for Processing Clinical Performance Ratings: Illustrative Applications in a Residency and Four Clerkships

Gerald J. Cason
Ture W. Schoultz
Carolyn L. Cason
Robert E. Glenn
Jerry G. Jones
Kimberly A. Golden
Nicholas P. Lang
K. Lee Lee Doyle

University of Arkansas for Medical Sciences

Innovations in Medical Education Exhibit
Association of American Medical Colleges
New Orleans, Louisiana
October 25-30, 1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. J. Cason

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM 870 109

A FLEXIBLE SYSTEM FOR PROCESSING CLINICAL PERFORMANCE RATINGS

Gerald J. Cason, Ph.D.
Office of Educational Development
SYSTEM DESIGNER & DEVELOPER

Ture W. Schoultz, Ph.D.
Associate Dean, College of Medicine
SYSTEM DEVELOPMENT SPONSOR

Carolyn L. Cason, R.N., Ph.D.
Graduate Program, College of Nursing
ASSOCIATE SYSTEM DESIGNER

APPLICATIONS in the COLLEGE of MEDICINE

Robert E. Glenn, M.D.
Program Director
PEDIATRIC RESIDENCY

JUNIOR YEAR CLERKSHIPS

Jerry G. Jones, M.D.
Clerkship Director
PEDIATRICS

Kimberly A. Golden, M.D.
Clerkship Director
SPECIALTIES

Nicholas P. Lang, M.D.
Clerkship Director
SURGERY

K. Lee Lee Doyle, Ph.D.
Clerkship Director
OBSTETRICS/GYNECOLOGY

UNIVERSITY of ARKANSAS for MEDICAL SCIENCES

Little Rock, AR 72205

Innovations in Medical Education Exhibit
Association of American Medical Colleges
New Orleans, Louisiana
October 25-30, 1986

Brief Abstract

The system collects, processes, and reports ratings of learners' performances in clinical settings. The system's 160 modular FORTRAN programs are presently run on a mainframe computer, but since the user interacts with an Assessment Scoring Coodinator not the computer, the user (e.g., course director) need have no technical knowledge of computers, printers, or scanners. The system maintains records across up to 30 separate assessment (rating and or testing) occasions within each user's course. Each user may specify up to 40 rating criteria per inventory and 30 inventories with different criteria. The rating criteria used in one course need have no relationship to those used in another. The paper first provides a detailed, generic description of the system followed by descriptions of its application in a Pediatrics Residency and Junior Year medical school clerkships (i.e., clinically oriented courses) in Pediatrics, Sub-specialties, Surgery, and Obstetrics/Gynecology. Each application section contains an example inventory, reports containing actual data, and the user's comments on the usefulness of the system. Our experience indicates that the system has greatly improved the quality, quantity, completeness, and timeliness of evaluation information available to educational administrators, faculty, and students alike. The system would likely prove useful in any training or practice setting, health-care or otherwise, where repeated ratings of human performance by multiple raters is needed to insure proper quality control.

A Flexible System for Processing Clinical Performance Ratings:
Abstract

Objective. Develop a system to collect, process, and report performance ratings of learners in practice settings that requires little faculty (observer) time and is both easy to use and flexible enough to encourage the acquisition of more timely, frequent, complete, and accurate rating data.

Method. The Performance Rating (PR) enhancement of the UAMS Objective Test Scoring and Performance Rating (OTS-PR) System became operational in 1981 after 4 years of development. Revisons and enhancement continue. OTS-PR is a set of modular FORTRAN programs run on a DEC-10. The user needs no technical knowledge of computers. Each user may specify up to: 40 items (criteria) per rating inventory, 30 different inventories, 200 subjects' (student, resident) identification numbers and names, and 30 assessment occasions (places or dates of tests, ratings). Machine scannable rating sheets are prepared by the computer's printer. Variable numbers of raters per subject are acceptable. OTS-PR maintains course level records across multiple assessment occasions. Example reports and analyses shown in the exhibit include: performance report for the individual student, including student's mean rating and standard error for item, category, and total scores; student reports for faculty use including rankings; and, inter-rater reliability.

Results. The PR system has been successfully used in a residency program and 4 Junior year clerkships. The clerkships processed both ratings and exam data. Users have been generally pleased. Both students and faculty report that the system is meeting or approaching its objectives. Quality and completeness of evaluative feedback to students have been improved by OTS-PR. The Dean's office is currently planning to make more direct use of PR reports from clerkships to prepare Dean's letters.

Conclusions. Voluntary adoption and continued use of OTS-PR to process clinical ratings of Pediatric residents and Junior year clerks in Pediatrics, Specialties, Surgery, and Obstetrics/Gynecology attests to the system's overall success in terms of flexibility, practicality, and utility.

Implications. In the future, systems similar to OTS-PR will alter the way medical education quality control is viewed and conducted. More appropriate, complex, and comprehensive data processing methods for analysis of evaluation information will fundamentally alter the typical medical teacher's approach to clinical performance evaluation. These new systems are essential to achieving the GPEP report's recommendations calling for a shift toward performance based assessment in medical education to prepare physicians for the 21st century.

A Flexible System for Processing Clinical Performance Ratings: The UAMS Performance Rating System*

The Performance Rating (PR) portion of the UAMS Objective Test Scoring and Performance Rating (OTS-PR) system was designed to help the clinical teacher evaluate students' clinical performance. The system has proven useful in a broad range of applications, including evaluating the clinical performance of medical students and residents, and the teaching performance of faculty. The system is especially useful when multiple raters provide ratings on each person rated and when records need to be kept across several assessments, e.g., ratings on different occasions or in different settings; or, when ratings are to be combined with scores from examinations. Because the performance rating system is an integral part of OTS-PR, scores from OTS-scored multiple-choice exams, hand-scored papers, and performance ratings can all be easily combined in the same computer files, i.e., "gradebook". The gradebook is a course level database containing subject (student) names and identification numbers linked to scores on multiple assessments (tests or ratings) designated by name and date. The system permits the user to specify weights (i.e., coefficients) to be applied to the results of each assessment in computing a final total or average combined score given in the gradebook reports.

In the following sections of this paper, first a more detailed, but generic description of the system is provided. This is followed by a separate section describing the application of the system to the assessment of clinical performance in a residency and four Junior year clerkships. Each of the application sections provides a brief description of the training program, setting, and evaluation procedures, including example rating inventories and reports containing actual data. All example rating sheets and reports included in this paper are photo-reduced to 65% of original full size. In all four of the clerkship application sections, the individual identified by the pseudonym Nancy C. Kildare is the same actual student who, in 1985-86, earned the ratings given in the example reports. Each application section concludes with the user's (i.e., Residency or Clerkship Director's) comments on the use and usefulness of the system. The final section addresses the issues related to making best use, in the Dean's office and in the residency application process (Dean's Letters), of the kinds of information provided by the Performance Rating portion of the OTS-PR System.

System Development

The original computer programs, developed beginning in 1975, contained only test scoring and course-level gradebook data-base capabilities. The system assumed that the faculty user would interact with the programs via computer terminal and operate the sheet scanner. At faculty insistence, the system was altered by the introduction of a Test Scoring Coordinator who functioned somewhat like the Assessment Scoring Coordinator as shown in Figure 1, but dealing only with test scoring. Thus, the user need have no technical knowledge of or skill in the use of computers, scanners, printers, etc.

*Prepared by Gerald J. Cason, Ph.D., System Designer and Developer and Carolyn L. Cason, R.N., Ph.D., Associate System Designer. Address communications concerning details to Gerald J. Cason, UAMS-OED-595, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 661-5720.

One aspect of the test scoring side of the system deserves mention because it still remains rare in the United States (G. Cason, 1982; 1986). The computer programs and scanner permit recording answers to a single test question as multiple marks on the test answer sheet. Also, scoring of such questions is on a partial credit rather than right-wrong basis. This permits routine scoring of what the National Board of Medical Examiners (NBME) calls multiple-true-false (MTF) Type X (i.e., make a separate mark for each correct alternative) rather than the more common, but less valid (Harasym, Norris & Lordsoneider, 1980) MTF Type K, which involves a coded, single mark response, e.g., "mark A if alternatives 1, 2 and 3 are correct."

Between 1977 and 1981, Gerald and Carolyn Cason collaborated in the design, testing, and implementation of first an experimental prototype rating information processing system; then, the original, operational OTS-PR rating programs and the General Purpose Rating sheet (G. Cason,1981). This original design and development project was conducted primarily under the sponsorship of Ture Schoultz and the College of Medicine's Council for Academic Affairs. However, the project also had the formal endorsement of an Associate Dean in each of the other three UAMS colleges, i.e., Nursing, Pharmacy, and Health Related Professions. During the same period, Gerald Cason completed a major redesign and upgrading of the test-scoring and course level data-base portion of the system. The Casons continue to design and develop system enhancements. Programmer Analyst Tom Lewis wrote first the experimental prototype computer programs then the original operational OTS-PR programs in the period 1977-81. Since 1981, Analysts Tom Hart and Chuck Desjardin have, with the assistance of Programmer Jalene Hawk, programmed most of the system's enhancements.

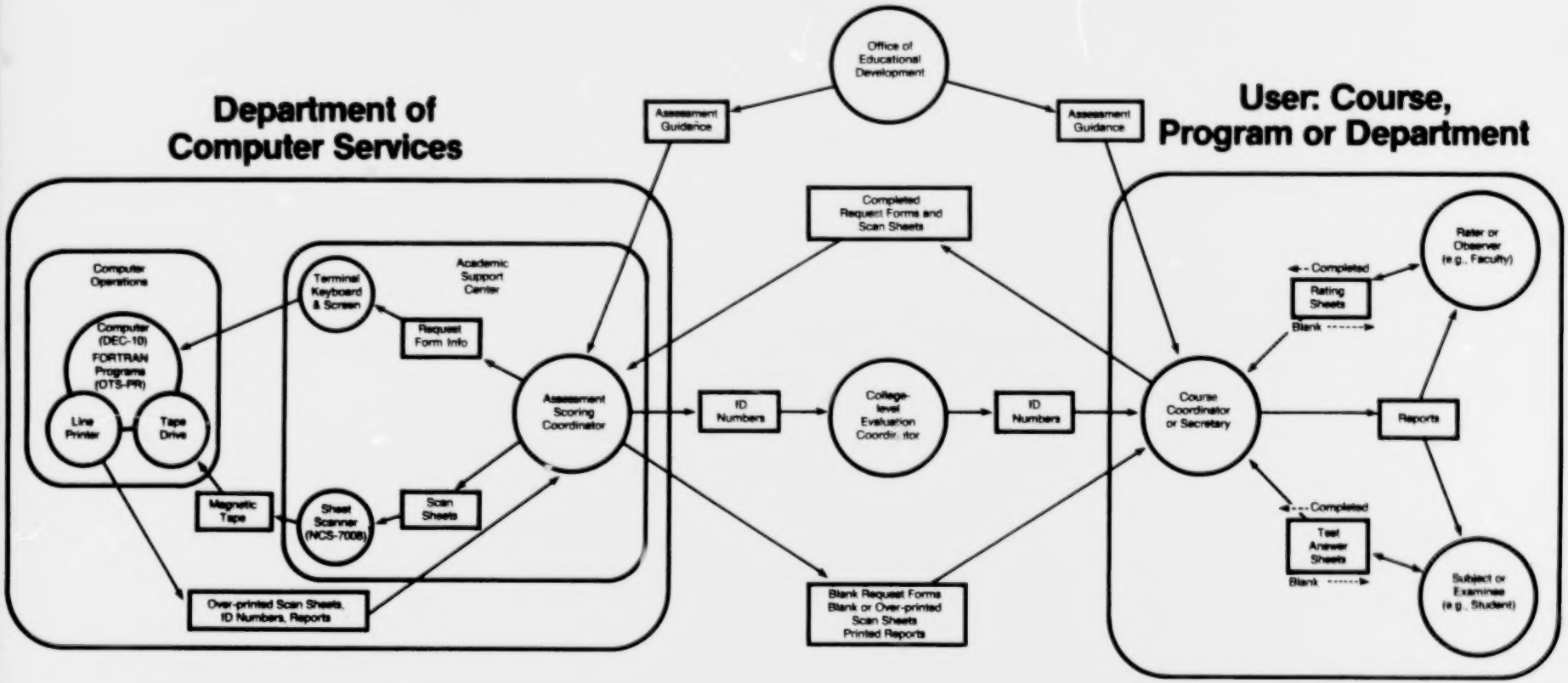
Hardware and Software

Currently, the OTS-PR System is a large, integrated set of modular, FORTRAN programs developed and run at UAMS on a Digital Equipment Corporation System 10 (DEC-10) mainframe computer. Table 1 lists some of the important, current size limitations of the system. All of these could be increased if necessary. OTS-PR assumes the use of a National Computer System (NCS) Trans-Optic 7008 scanner with UAMS developed DOSSIER computer programs for scanner control and data pre-processing by the scanner's micro-computer. As illustrated in Figure 1, data are transferred from NCS Trans-Optic scan sheets to magnetic tape by the NCS-7008 and DOSSIER programs; then, from tape to DEC-10 disk for processing by the OTS-PR FORTRAN programs. All reports are output on a standard (800-1200 lines per minute) line printer. Both a detailed Test Scoring Coordinator's guide (58 pages) covering scanner operation, data transfer and processing, and system management procedures; and, a detailed OTS-PR user's guide (59 pages) are available.

Table 1. Size Limitations of the OTS-PR System

Feature	Maxima
Subject (e.g., student) names and ID numbers	200
Rater (e.g., faculty) names and ID numbers	200
Assessment occasions (tests or ratings) per course	30
Categories (sub-scales) per assessment	12
Rating inventories (item sets) per course	30
Rating items (criteria) per rating inventory	40
Test items (questions) per test	240

Information Flow in the System



8

9

Ex 1, page 7

Operation of the System

Figure 1 provides an overview of information flow in the system. At the left-center is the Assessment Scoring Coordinator (ASC) in the Department of Computer Service's Academic Support Center. The system user, at the right-center in Figure 1, interacts with the ASC, who in turn interacts with the scanner and computer. Information is passed from the user to the ASC in two modes: via request forms and scan (i.e., machine scannable data) sheets. Table 2 provides a descriptive listing of the request forms and scan sheets used in the system.

Table 2. Request Forms and Scan Sheets Used in the OTS-PR System

Manual Input via Request Forms
- New User (Start-up)
- Report and Gradebook Selection
- Rating Inventory (Scale, Item Text) Specification
- Over-printing Scan Sheets
- Gradebook Weight Change
- Enrollment Change
- Assessment Scoring

Automated Input via Scan Sheets
- Student Enrollment
- Rating (Must be over-printed)
- Test Answer (May be over-printed)
- Control (Must be over-printed)

The faculty user first requests that the ASC set-up an OTS-PR Course (i.e., file in a highly secured area of the computer) for his/her use. This is done by the user either filling out a New User request form, or providing the same information orally to the ASC (who then completes the request form for the user). The user supplies the subjects' names (i.e., the people or things to be rated) entered on machine scannable Enrollment Sheets. In the College of Medicine the enrollment scan sheets are used only once, i.e., when a Freshman or transfer student enters the program. Thereafter, names and ID numbers are copied from master rosters kept in disk storage by the ASC. Standard Inventory Specification forms are provided to the user to communicate the inventory to the ASC. The inventory includes the definition of the five point scale and up to 40 Likert-type items (criteria) on which subjects are to be rated. Inventories are entered via terminal keyboard (or if previously used in the system, copied from disk or tape into the current user's area) by the ASC.

Figure 2 provides an illustration of a rating scan sheet and a hypothetical inventory. The computer's line printer over-prints both identifying information and the inventory text on the machine scannable sheets. This includes "slugging" identification data in the scannable data grids. Using the line printer to print both the inventory and subject identifying data on the rating sheet provides the user maximum flexibility, easy editing and revision, while also minimizing the quantity of information manually entered on the sheets. As shown in Figure 2, the rater need record only his or her ratings. The observer or rater records a rating of 1, 2, 3, 4, 5, (5 is always best), "no opinion", or "not applicable", by marking a circle next to each criterion. Written comments by the rater may be recorded on the back of the rating sheet. Optionally, an ID number uniquely identifying the rater may be manually entered. This permits the

generation of reports on the performance of each rater. (This option is currently, 1986-87, in restricted use in our Pediatric Residency and Junior Clerkship.) Otherwise, a default number is printed in the rater ID grid and processing proceeds on the assumption that each sheet comes from an independent, but anonymous rater. The subject's ID number used on the rating sheet is not confidential. A different, confidential number is used in reports and records.

While the basic design of the processing programs and rating sheets assumes brief (i.e., one criterion per line; 35 characters per descriptor) items will be used, more extended definitions are possible in either of two ways. A separately produced assessment guide (such as that used in our Pediatric Residency) may be provided to the rater which contains more nearly complete definitions of the criteria briefly designated on the rating sheet. Alternatively, the category function may be used.

The system permits defining categories (i.e., explicitly defined sub-scales) of items. On the rating sheet, a category begins with a category name. This is followed on the rating sheet by all the items in that category. A category score is computed as a weighted average of the ratings given to all the items in that category. Although not designed for the purpose, the user can use this feature to print a multi-line definition of a criterion on the rating sheet; as is illustrated in the Surgery Clerkship section of this paper. To accomplish this, one of the items within the category is called something like, "Overall rating for category N," and given all the points associated with that category. The rater records a rating on the sheet only for the "overall" item in each category. The other lines in the category are not functional items but part of the definition of the criterion. However, each of these lines are "items" in the system's data processing logic and must, therefore, be assigned a point value of zero. In the resulting reports, the category scores give the substantive information on each rating criterion. The reports of item results are meaningless, because the "items" are only fragments of definitions of criteria. This leads to a noticeable messiness in the appearance of the item-level reports.

Up to 30 different inventories may be used within a single OTS-ER Course. Each inventory in a course has a unique ID (item set) number which is printed on the rating sheet in scannable form. When the scanner is processing rating sheets, a single inventory ID number is specified. Any sheet not having that number is mechanically ejected by the scanner into a separate bin. This prevents data of different types or from different sources from being inadvertently mixed together. For example, if both a resident and an attending physician rate the student, but separate reports from each kind of rater are required, different inventory numbers are assigned to the sheets to be used by each rater group, even if the substantive criteria used by each group are identical. All our clerkship applications make some use of this feature. One potential disadvantage of segmenting data in this way is the system's inability to compute inter-rater agreement (reliability) statistics when the explicit number of raters per student within a data set is less than two. Both a loss in information (i.e., reliability and validity of the evaluation results) and inability to determine inter-rater agreement result when one person (commonly the attending physician) is allowed to summarize, interpret, and report (on a single rating sheet) information obtained from several others.

Field tests with prototype rating data processing programs (circa 1978-80) demonstrated that it is essential to keep to a bare minimum the quantity of data

manually entered in the scannable grids. Otherwise, errors increase and excessive time for recording reduces the acceptability and usefulness of the system. For this reason, in many cases where rater ID numbers are used, it is best if a clerk enters these rather than the rater.

The user specifies (on the Printing Request Form) how many copies of the rating sheet are needed per subject, i.e., person to be rated. In student performance evaluation, the subjects are the students that are to be evaluated and the number of rating sheets produced is equal to the number of attending physicians, residents, or other judges that will be evaluating each student. The ASC interacts with the OTS-ER programs via computer terminal to set up the requested scan (i.e., rating) sheet printing job. A member of the Computer Operations staff mounts a special high density carbon ribbon on the printer, loads, and aligns the continuous-form rating sheets in the line printer, then executes the print job. The operations staff bursts (i.e., separates) the continuous-form into individual rating sheets and removes the tractor strips from the edges of the forms prior to sending the sheets to the ASC for transmital to the user.

The rating sheets are passed from the ASC to the Course Director or course-level evaluation coordinator (usually the program, department, or course director's secretary) for distribution to the raters. When the rating sheets are completed, the evaluation coordinator collects them into a pile, places a completed Assessment Scoring request on top as a cover-sheet, and gives the batch to the ASC for processing.

As suggested in Figure 1, the ASC enters some information from the Scoring Request form (e.g., assessment name, date, inventory number, etc.) via terminal keyboard and feeds the rating sheets through the scanner. The data from the scan sheets are transferred to a magnetic tape. The tape is mounted on the DEC-10's tape drive and the rating data are tranferred to disk. The rating scoring job is run in batch mode based on information in the user's course file (e.g., subject names and ID numbers, requested reports, etc.), the Scoring Request and data from the rating sheets. The results are computer printed reports summarizing the performance of each subject rated, reports on the behavior of the inventory, i.e., inter-rater agreement (reliability), average on each item, etc. and potentially reports on the performance of individual raters. Table 3 provides a partial listing of the seventeen reports currently available to the user.

The reports obtained by a particular user are determined by the user's selection from the menu of reports on the Report Selection and Gradebook Selection request forms. Selection of which gradebook to receive reflects the user's choice of arithmetic to be used in computing a total or average across assessments. The raw score gradebook provides reports which give the sum of the points across assessments as the cumulative score for a student. The weighting in the raw score gradebook is implicit, i.e., the relative magnitudes of the maximum points on various assessments is assumed to reflect their relative importance or value. The percent score and Z-score gradebooks permit explicit weighting. As the names imply, the original measures (scores) in these two gradebooks are represented respectively as either percent scores or Z-scores. In the percent score gradebook (used in our Surgery Clerkship), the over-all cumulative performance of a student is represented by a weighted average of percent scores. In the Z-score gradebook (used in our Pediatric, Specialties, and Ob/Gyn Clerkships), original scores are represented by Z-scores (i.e.,

standard scores with a mean of 500 and a standard deviation of 100). Cumulative performance is expressed as the Z-tranformed weighted average of Z-scores on the original assessments. (An average or mean of Z-scores is not likely to be a Z-score; so, the averages of Z's are converted back to the Z-scale in the Z-score gradebook.) These three approaches to combining scores from different assessments into a single cumulative score can and usually do lead to different student rankings and score differences between students.

Table 3. Partial Listing of the Reports Available from the OTS-PR System

For Current Assessment (Test or Rating)
- Assessment Instrument Analysis
  - Analysis Summary
  - Item Analysis
- Subject (e.g., Student) Performance
  - Department (detailed scores for archive)
  - Subjects by Rank Order
  - Histogram of Subject Scores
  - Posting (Subjects' names excluded)
  - Individual Subject Performance
- Individual Rater (e.g., Faculty) Performance

Across Multiple Assessments (Tests or Ratings)
- For Course Coordinator's Use
  - Alpha-ordered, all subjects' scores to date
  - Rank ordered, subjects' cumulative weighted totals
- Subject's Individual Cumulative Scores
- Posting (Subjects' names excluded)
  - ID-ordered all scores to date
  - ID-ordered weighted total

Example Reports

Of the 17 available reports, examples of only four will be provided in this section. These were chosen to provide the most information in the least space about the performance rating portion of the system. Although the reports are structured to be internally consistent with each other and the rating inventory in Figure 2, they contain only synthetic data and fictitious subject names. The fictitious student's name, Noel C. Kildare, appears on the the example rating sheet and in each of the example reports in this section to assist in showing what information is redundant between reports.

The Rating Analysis Summary Report is illustrated in Figure 3. It provides information on the performance of the assessment inventory and procedure. Information is given at the category (sub-scale) and total score level. If two or more raters rated each subject, meaningful inter-rater reliabilities are reported. (When only one rater rates each subject such as in the example reports from our Surgery and Ob/Gyn Clerkships, a meaningless zero is reported for the reliability.) As the number of raters rating each subject (student) may vary, the reliability of the average (i.e., geometric mean) number of raters rating the average subject is reported. It should be noted that the single rater reliability is the same as the single rater convergent validity (Cason & Cason, 1984; Cason, Cason, & Stritter, 1986), but validity expands more slowly with increased independent observations than does reliability. More detailed information on each rating item, such as the distribution of ratings given, is

Figure 2. Generic Rating Sheet and Inventory

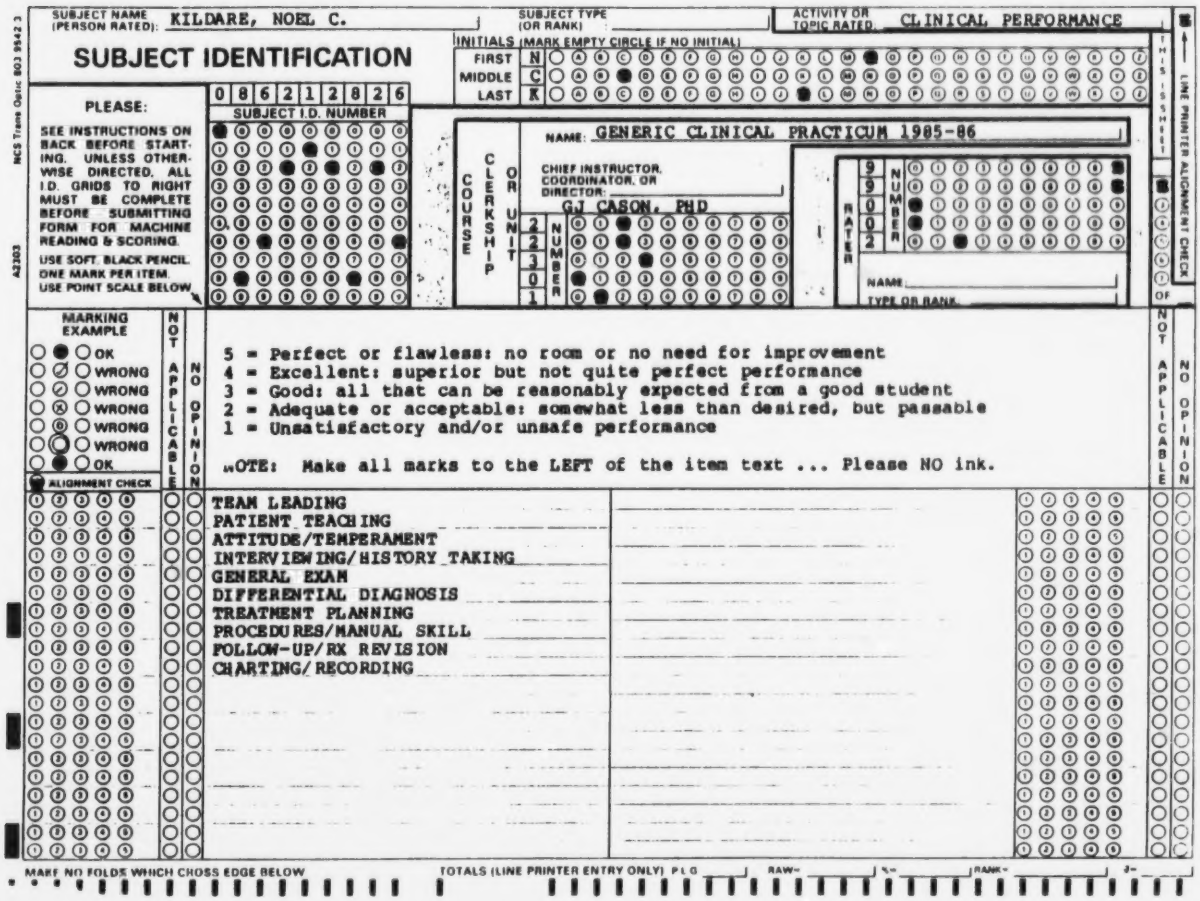

SUBJECT NAME (PERSON RATED): KILDARE, NOEL C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED: CLINICAL PERFORMANCE
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING. UNLESS OTHERWISE DIRECTED, ALL I.D. GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT, BLACK PENCIL. ONE MARK PER ITEM. USE POINT SCALE BELOW.
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
COURSE OR CLERKSHIP OR UNIT NUMBER 2 2 3 0 1
NAME: GENERIC CLINICAL PRACTICUM 1985-86
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR: GJ CASON, PHD
RATER NUMBER 9 9 0 0 2
NAME:
TYPE OR RANK:
THIS IS SHEET 1 OF 1
LINE PRINTER ALIGNMENT CHECK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsafe performance
NOTE: Make all marks to the LEFT of the item text ... Please NO ink.
TEAM LEADING
PATIENT TEACHING
ATTITUDE/TEMPERMENT
INTERVIEWING/HISTORY TAKING
GENERAL EXAM
DIFFERENTIAL DIAGNOSIS
TREATMENT PLANNING
PROCEDURES/MANUAL SKILL
FOLLOW-UP/RX REVISION
CHARTING/RECORDING
MAKE NO FOLDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) PLG RAW= %= RANK= Z=

Figure 3. Generic Rating Analysis Summary

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 29-Jul-86 14:40 by the UAMS OIS/PR System (version B0) as implemented at UAMS

Test: 3 - INPATIENT RATING - FACULTY
Instructor: GJ CASON, PHD
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Dept: EDUCATIONAL DEVELOPMENT
Slot: 595 Phone:661-5720
Subjects rated: 60 Absent: 68 Withdrawn: 8

	Category	N Rated Items & Total Points (1)		Average Score Raw	5 Pt	1 Rater Reliab.	Mean N Raters & Reliability (2)		Standard Dev Raw	5 Pt	Std Error of Measure Raw	5 Pt	Z (3)
1	TEAM LEADING	1	5	4.0	3.97	0.17	3.7	0.43	0.4	.428	0.32	.32	75
2	PATIENT TEACHING	1	5	3.8	3.80	0.21	3.7	0.49	0.5	.513	0.37	.37	71
3	ATTITUDE/TEMPERMENT	1	5	4.1	4.08	0.20	3.7	0.48	0.5	.455	0.33	.33	72
4	INTERVIEWING/HISTORY TAKING	1	5	3.9	3.92	0.21	3.5	0.48	0.5	.540	0.39	.39	72
5	GENERAL EXAM	1	5	3.7	3.67	0.36	2.7	0.60	0.6	.552	0.35	.35	63
6	DIFFERENTIAL DIAGNOSIS	1	5	3.6	3.63	0.18	3.2	0.41	0.5	.481	0.37	.37	77
7	TREATMENT PLANNING	1	5	3.6	3.61	0.21	3.6	0.49	0.4	.435	0.31	.31	72
8	PROCEDURES/MANUAL SKILL	1	5	3.6	3.64	0.12	2.1	0.22	0.5	.486	0.43	.43	88
9	FOLLOW-UP/RX REVISION	1	5	3.6	3.64	0.35	2.8	0.60	0.6	.555	0.35	.35	63
10	CHARTING/RECORDING	1	5	3.6	3.57	0.24	3.0	0.49	0.5	.494	0.35	.35	72
	Overall	10	50	37.4	3.74	0.30	3.8	0.61	4.0	.398	2.47	.25	62

Figure 4. Generic Students by Rank Report

STUDENTS BY RANK (Current Rating)

Prepared 30-Jul-86 14:27 by the UAMS OIS/PR System (version B0) as implemented at UAMS

Test: 3 - INPATIENT RATING - FACULTY
Instructor: GJ CASON, PHD
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Dept: EDUCATIONAL DEVELOPMENT
Slot: 595 Phone:661-5720
Subjects rated: 60 Absent: 68 Withdrawn: 8

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	N OF RATERS	NAME
	Averages	37.41	3.74			500	3.8	(Fictitious names and data)
1	286001297	49.33	4.93	1	98	799	3	ZEIMANOVICH, MARINA K.
2	286007980	46.22	4.62	2	96	721	3	BEAKER, JULIO M.
18	286002197	39.17	3.92	18	70	544	4	KILDARE, NOEL C.
60	286001645	28.67	2.87	60	0	281	3	ARNOLD, FRANKLIN D.

Figure 5. Generic Individual Performance Report

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 29-Jul-86 14:53 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: NOEL C. KILDARE
From: GJ CASON, PHD
Re: Rating/test 3-INPATIENT RATING - FACULTY

Dept: EDUCATIONAL DEVELOPMENT
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Item	5 Point Scale 1.....2.....3.....4.....5	5 Pt Score: Your Mean=x	Your SEM	Class Mean=c	Class StdDev	Raw Score: Perfect p	Raw Score: Yours xp/5	N of Raters
Class Overall Mean Rating = 3.74-->	C							
Your Overall Mean Rating = 3.92-->	X							
1 TEAM LEADING	.-.x-	4.00	.32	3.97	.428	5	4.00	4
2 PATIENT TEACHING	.x.-	3.75	.36	3.80	.513	5	3.75	4
3 ATTITUDE/TEMPERMENT	.-.-C---x-	4.75	.32	4.08	.455	5	4.75	4
4 INTERVIEWING/HISTORY TAKING	.-C---x-	4.50	.38	3.92	.540	5	4.50	4
5 GENERAL EXAM	x---C-.-	3.00	.38	3.67	.552	5	3.00	2
6 DIFFERENTIAL DIAGNOSIS	x-C.-.-	3.33	.37	3.63	.481	5	3.33	3
7 TREATMENT PLANNING	xC.-.-	3.50	.30	3.61	.435	5	3.50	4
8 PROCEDURES/MANUAL SKILL	Cx-.-	3.67	.41	3.64	.486	5	3.67	3
9 FOLLOW-UP/RX REVISION	xC.-.-	3.50	.39	3.64	.555	5	3.50	2
10 CHARTING/RECORDING	x---C.-.-	3.00	.39	3.57	.494	5	3.00	2

Rating Scale
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsafe performance

Definition of Symbols
C = class overall 5 pt score: 3.74 StdDev: .398
X = your overall 5 pt score: 3.92 SEM : .24
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 39.17 (out of perfect 50) yields: 78.3% Z= 544 Rank= 18 (out of 60). Class ave raw score 37.41

Figure 6. Generic Individual Cumulative Score Report

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 29-Jul-86 15:03 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: NOEL C. KILDARE
From: GJ CASON, PHD
Re: Your test scores and cumulative scores to date

Dept: EDUCATIONAL DEVELOPMENT
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
1	38.72	58	19-Jun-86	OUTPATIENT RATING - SUPERVISOR	1	44.50	45.00	98.89	4.94	3	624
2	34.84	61	19-Jun-86	EMERGENCY ROOM RATING - STAFF	1	ABSENT	45.00				
3	37.41	60	29-Jul-86	INPATIENT RATING - FACULTY	1	39.17	50.00	78.33	3.92	18	544
4	50.30	121	29-Jul-86	COMPREHENSIVE EXAMINATION	3	30.50	100.00	30.50	1.53	116	298
	Currently Enrolled:128			Total applicable weights:	5						
				Your weighted normalized scores, and derived rank and Z:				53.74	2.69	105	403
				The weighted mean of your Z scores is 412.4, derived rank and Z:						110	374

provided in the Rating Item Analysis Report (not shown). The rating analysis information provided in these reports can assist the user to improve both the reliability and validity of his or her inventory through judicious editing, revision, and practical trial. At UAMS the Office of Educational Development provides assistance to the user in making best use of this kind of information. Some of the inventories used in the residency and clerkship programs reported below have been improved in this manner. These reports permit the user to constantly monitor the technical quality of the assessment inventories and procedures used in his or her course.

The Students by Rank Order report illustrated in Figure 4 provides information about all members of a class. It provides a listing of all students, rank ordered from highest-scoring to lowest-scoring with the total score reported in several different units, e.g., raw, percent, rank, and Z-score. Also given is the number of raters or rating sheets the total scores were averaged across for each student. As shown in Figure 4, the number of raters rating each student need not be the same. Because the students' names are given along with their confidential ID numbers, this report is intended for faculty use only. The Posting Reports (not shown) contain the same kind of information, but arranged in ID order and omitting student names. As the names of the posting reports suggest, they are intended to communicate performance results to the students.

The Individual Performance Report (IPR) illustrated in Figure 5 provides information to or about a single individual regarding a single rating occasion. This is the only report which gives item-level rating information on the performance of an individual subject. It gives item-level average ratings in both tabular and graphic form. The "x" profile makes it easy to rapidly determine, by visual inspection, this subject's relative strengths and weaknesses. The "c" profile provides a comparison with the average ratings obtained by all the members of the group on whom data were included when the report was prepared. The standard error of measurement (at item, category, and total score levels) is computed based on the number of raters rating this individual. Performance of all members of the class at category and total score levels is given in the Department Report (not shown). The Department Report like the Student by Rank Report is intended for faculty use. When rater ID numbers are used, the report on an individual rater is similar in structure to that shown in Figure 5. However, the rater report gives the average rating that the rater assigned to the subjects he or she rated.

The Individual Cumulative Score Report (ICSR) illustrated in Figure 6 provides information to or about a single individual. It lists all the assessments to date, giving the name and date of each, the class average, this individual's score on each and weighted totals computed in alternative ways and expressed in several units. Cumulative information on all members of the class is given in the various Gradebook reports (not shown). The example ICSR in the Surgery Clerkship section of this paper illustrates how absences and weighting an assessment zero are treated.

A Flexible System for Processing Clinical Performance Ratings: Application in a Pediatric Residency*

The Pediatric Residency Program at the University of Arkansas for Medical Sciences (UAMS) is located primarily at Arkansas Children's Hospital (ACH). ACH is the only pediatric hospital in the state. ACH is, therefore, the state-wide referral center for all pediatric teritary care in the state. ACH also serves as primary care provider for a part of the Little Rock population. The nospital consists of 195 beds; 30 NICU beds, 12 PICU beds and 12 CVICU beds. There is a 24 hour emergency room which is used by 25,000 patients yearly. The General Pediatric Clinic sees another 20,000 patients. There are over 40 other clinics which together account for 90,000 ambulatory visits yearly. There are 60 pediatric subspecialists on the faculty and about 12 pediatric surgical subspecialists. All of these pediatric physicians and surgeons are full time teaching faculty. This insures that our residents receive a wide range of teaching input and therefore a wide range of evaluations.

The pediatric residency is three years long. There are 13 residents accepted each year at each of the three levels of training. The program fills its quota yearly. There is also a four year medicine-pediatrics program with three positions at each level. Although this program also fills its quota, currently there are only eight med/peds residents because of attrition.

Our program is rather easy to describe. The resident's first year of training is spent as an intern in basic pediatrics, i.e., approximately 50% inpatient, 25% outpatient and 25% nursery. The resident's second year is spent primarily with the subspecialty services, working in the specific clinics and acting as consultant on inpatients. There is also a small amount of time spent in supervisory positions in the nursery and inpatient rotations. The third year resident returns to basic pediatrics as a supervisor of the first year resident.

Performance Evaluation Procedures

Since all rotations in our residency program are monthly, the residents' performance evaluations are completed on a monthly basis. The rating inventory is completed by the faculty person who is directly responsible for a resident. The rating sheet (Figure 8) has limited space for describing individual criteria in the inventory, therefore, a supplementary rating guide (see Figure 7) is sent out with each rating sheet. The guide defines in more detail the scope of the inventory and the meaning of each criterion listed on the rating sheet. Although we consider evaluations of the resident by the faculty our most important inventory, other inventories are also in use. We ask the residents to evaluate themselves (monthly), the teachers (monthly), and the services (yearly). The evaluation of the resident by faculty is the only one insisted on by the departmental chairman. The others are encouraged but are completed voluntarily.

*Prepared by Robert E. Glenn, M.D., Program Director. Address communications concerning details of this application to this author at UAMS-COM-512, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 370-1874 (Arkansas Children's Hospital).

Figure 7. Guide 1: Clinical and Ward Performance Assessment*

Check the SPECIAL NOTE to make certain you have a computer prepared rating sheet intended for the observer group to which you actually belong. Make sure you have a rating sheet prepared for the correct person. Do not attempt to alter name or ID numbers on rating sheets. If the sheet you have is inappropriate, get a replacement from the Program Director's secretary or as directed.

On a 1 to 5 scale (5 is best) using definitions below, rate clinical and ward performance observed. Record ratings in number 2 pencil. Mark your rating or other response to the LEFT of each criterion printed on the rating sheet. If you have inadequate information to rate a criterion, mark "no opinion". If a criterion does not apply to the circumstances of the observed performance, mark "not applicable". Items omitted (left blank) scored as if "no opinion" were marked. On the back of the rating sheet explain ratings of 1 and 5, make comments, date and sign. Please do not fold or staple rating sheet. Return rating sheet to Pediatric Residency Program Director's Secretary.

(* Adapted by Gerald J. Cason, Ph.D., Office of Educational Development, University of Arkansas for Medical Sciences, November 1984, from American Board of Pediatrics' resident evaluation form. Criteria B - H correspond to A - G on the Board's form. Preparation supported in part by USPHS General Internal Medicine and General Pediatrics Division Grant Number 5D28PE16028-02.)

All items pertain to effectiveness of Resident's (Clinician's) behavior.

A. PREVENTION, EARLY INTERVENTION, WHOLE-CHILD FOCUS

INEFFECTIVE (1): 1: Ignores pt education. Fails to assess developmental milestones. Fails to use literature and/or other Health Team Personnel in pt care and education. Does not use whole-child concept: fragmented approach lacks integration or perspective. Focuses on disease or injury in isolation.

EFFECTIVE (5): 5: Uses anticipatory guidance with respect to nutrition, safety, immunizations, growth and developmental milestones. Uses whole-child concept in integrated, comprehensive approach.

B. GATHERING DATA BY HISTORY

INEFFECTIVE (1): 1: Directs interview toward preconcieved impressions. Fails to include social, family, and developmental history routinely. Accepts facts as recieved. Keeps disorganized, unreliable records.

EFFECTIVE (5): 5: Establishes rapport, defines roles and purposes clearly. Uses open-ended questions. Adapts vocabulary to socio-cultural and educational status of historian. Recognizes and responds to nonverbal and attitudinal cues.

C. GATHERING DATA BY PHYSICAL EXAM

INEFFECTIVE (1): 1: Alarms patients by clumsy or abrupt approach. Shows little concern for patients' comfort. Performs careless or incomplete exams. Fails to verify findings.

EFFECTIVE (5): 5: Shows judgment as to appropriate attention to a given area or manipulation. Interprets physical findings accurately. Repeats exam when findings are equivocal or obscure.

D. GATHERING DATA BY LAB STUDIES (INCL. RADIOGRAPHIC)

INEFFECTIVE (1): 1: Uses lab indiscriminately. Is inconsiderate of patients' comfort in selecting lab studies. Is unaware of influence of timing, sequence on lab studies. Interprets results incorrectly. Technically clumsy.

EFFECTIVE (5): 5: Selects invasive methods only when essential. Considers cost in selecting lab studies. Interprets lab results correctly. Preforms technical diagnostic procedures skillfully.

E. ASSESSING DATA AND ARRIVING AT DIAGNOSIS

INEFFECTIVE (1): 1: Ignores part of data in formulating diagnosis. Avoids reaching conclusion despite a reasonable data base. States problem in incorrect or confusing terminology.

EFFECTIVE (5): 5: Formulates sound diagnostic hypotheses. Is specific in problem definition. Explains working diagnosis to family appropriately.

F. MAINTAINING HEALTH AND MANAGING PROBLEMS

INEFFECTIVE (1): 1: Performs technical therapeutic procedures poorly Is slow to recognize emergency situations. Persists in original course despite indications to change. Rejects or ignores importance of patient education. Ignores psycho-social aspects of patient care.

EFFECTIVE (5): 5: Performs technical therapeutic procedures skillfully. Is prompt to respond to emergencies. Is effective and efficient in choosing course of management. Keeps assessment of patient and management plan current. Explains management plans to patients and parents effectively. Uses anticipatory guidance appropriately.

G. INTERPERSONAL RELATIONS WITH OTHER MEMBERS OF HEALTH TEAM

INEFFECTIVE (1): 1: Does not request assistance when needed. Fails to recognise and use professional skills of others. Is unable to give or take advise gracefully.

EFFECTIVE (5): 5: Utilizes contributions of others in total patient care and education. Supervises effectively. Teaches effectively.

H. WORK HABITS AND PERSONAL QUALITIES

INEFFECTIVE (1): 1: Does not accept responsibility. Is unaware of own limits, denies mistakes. Demonstrates personal dishonesty. Is unable to cope with personal or professional problems. Fails to preserve the confidentiality of medical information.

EFFECTIVE (5): 5: Demonstrates a caring attitude toward patients and families. Demonstrates personal honesty. Seeks to update knowledge in the interest of current patients' problems. Demonstrates appropriate initiative, energy, commitment. Makes decisions promptly and appropriately when risk of uncertainty or error must be accepted.

File:GUID1.FIL[2156,2156] GJC:17-MAY-85

Figure 8. Rating Sheet and Inventory from a Pediatric Residency

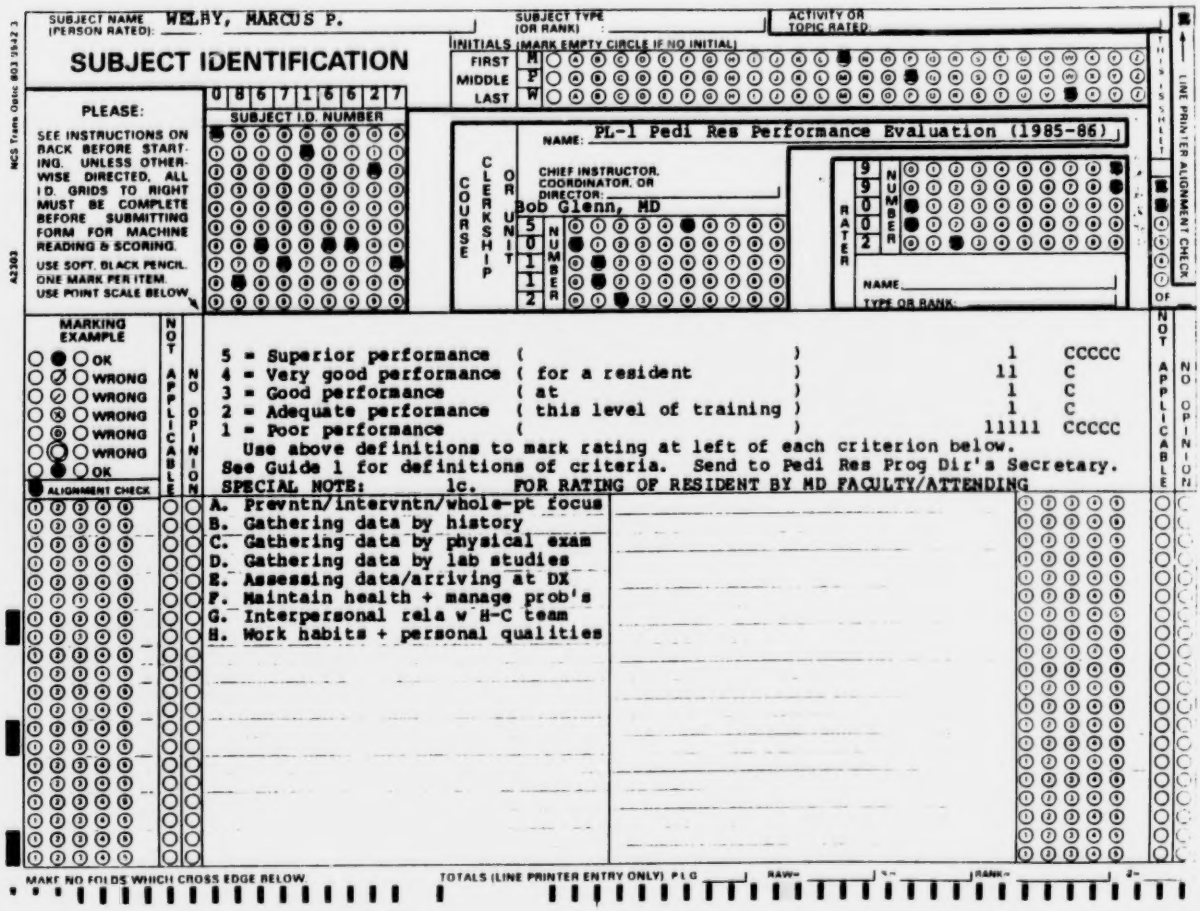

SUBJECT NAME (PERSON RATED): WELBY, MARCUS P.
SUBJECT TYPE (OR RANK):
ACTIVITY OR TOPIC RATED
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST M
MIDDLE P
LAST W
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING. UNLESS OTHERWISE DIRECTED, ALL I.D. GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT, BLACK PENCIL. ONE MARK PER ITEM. USE POINT SCALE BELOW.
0 8 6 7 1 5 6 2 7
SUBJECT I.D. NUMBER
NAME: PL-1 Pedi Res Performance Evaluation (1985-86)
COURSE
CLERKSHIP OR UNIT
5 0 1 1 2
NUMBER
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR:
Bob Glenn, MD
RATER
9 9 0 0 2
NUMBER
NAME
TYPE OR RANK
LINE PRINTER ALIGNMENT CHECK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Superior performance ( ) 1 CCCCC
4 = Very good performance (for a resident) 11 C
3 = Good performance (at) 1 C
2 = Adequate performance (this level of training) 1 C
1 = Poor performance ( ) 11111 CCCCC
Use above definitions to mark rating at left of each criterion below.
See Guide 1 for definitions of criteria. Send to Pedi Res Prog Dir's Secretary.
SPECIAL NOTE: 1c. FOR RATING OF RESIDENT BY MD FACULTY/ATTENDING
A. Prevntn/intervntn/whole-pt focus
B. Gathering data by history
C. Gathering data by physical exam
D. Gathering data by lab studies
E. Assessing data/arriving at DX
F. Maintain health + manage prob's
G. Interpersonal rela w H-C team
H. Work habits + personal qualities
MAKE NO FOLDS WHICH CROSS EDGE BELOW.
TOTALS (LINE PRINTER ENTRY ONLY) PLG RAW= RANK=

Figure 9. Rating Analysis Summary from a Pediatric Residency

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 5-Aug-86 14:22 by the UAMS OTS/PR System (version B0) as implemented at UAMS

Test: 1 - FACULTY RATING OF RESIDENT
Instructor: Bob Glenn, MD
Course: PL-1 Pedi Res Performance Evaluation (1985-86)

Dept: Pediatrics
Slot: 512 Phone:370-1016
Subjects rated: 14 Absent: 0 Withdrawn: 1

	Category	# Rated Items (1)	Total Points (1)	Average Score Raw	Average Score 5 Pt	1 Rater Reliab.	Mean # Raters (2)	Reliability (2)	Standard Dev Raw	Standard Dev 5 Pt	Std Error of Measure Raw	Std Error of Measure 5 Pt	Measure Z (3)
1	A. Prevntn/intervntn/whole-pt focus	1	5	3.6	3.59	0.08	9.1	0.45	0.3	.308	0.23	.23	74
2	B. Gathering data by history	1	5	3.5	3.49	0.14	9.0	0.60	0.3	.332	0.21	.21	63
3	C. Gathering data by physical exam	1	5	3.5	3.55	0.11	9.0	0.52	0.3	.309	0.22	.22	70
4	D. Gathering data by lab studies	1	5	3.5	3.48	0.04	9.2	0.25	0.2	.229	0.20	.20	87
5	E. Assessing data/arriving at DX	1	5	3.5	3.53	0.11	9.0	0.53	0.4	.351	0.24	.24	68
6	F. Maintain health + manage prob's	1	5	3.6	3.63	0.09	8.9	0.48	0.3	.322	0.23	.23	72
7	G. Interpersonal rela w H-C team	1	5	3.8	3.79	0.07	9.2	0.41	0.3	.306	0.23	.23	77
8	H. Work habits + personal qualities	1	5	3.9	3.94	0.00	9.1	0.01	0.2	.231	0.23	.23	100
	Overall	8	40	29.1	3.63	0.10	9.2	0.51	2.1	.266	1.49	.19	70

Figure 10. Students by Rank Report from a Pediatric Residency

STUDENTS BY RANK (Current Rating)

Prepared 5-Aug-86 14:22 by the UAMS OIS/PR System (version B0) as implemented at UAMS — Page 1

Test: 1 - FACULTY RATING OF RESIDENT
Instructor: Bob Glenn, MD
Course: PL-1 Pedi Res Performance Evaluation (1985-86)

Dept: Pediatrics
Slot: 512 Phone:370-1016
Subjects rated: 14 Absent: 0 Withdrawn: 1

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	# OF RATERS	NAME
	Averages	29.07	3.63			500	9.2	(Actual data; fictitious names)
1	585000359	32.36	4.05	1	92	654	11	AJGC
2	585000334	32.30	4.04	2	85	651	10	BLSR
3	585000235	31.30	3.91	3	78	605	6	WELBY, MARCUS P.
4	585000086	31.00	3.88	4	71	590	9	GJNF
5	585000482	29.69	3.71	5	64	529	8	DWCS
6	585000417	29.40	3.68	6	57	515	10	DDBJ
7	585000011	29.22	3.65	7	50	507	9	BGCM
8	585000466	28.97	3.62	8	42	495	11	FCHK
9	585000441	28.57	3.57	9	35	476	7	GDMS
10	585000516	27.78	3.47	10	21	439	11	HENW
11	585000458	27.78	3.47	10	21	439	11	IKMV
12	585000052	26.50	3.31	12	14	379	8	JWSS
13	585000276	26.49	3.31	13	7	379	10	KKAC
14	585000490	25.66	3.21	14	0	340	8	LAHK

Figure 11. Individual Performance Report from a Pediatric Residency

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 5-Aug-86 14:23 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: MARCUS P. WELBY
From: Bob Glenn, MD
Re: Rating/test 1-FACULTY RATING OF RESIDENT

Dept: Pediatrics
Course: PL-1 Pedi Res Performance Evaluation (1985-86)

Item — 5 Point Scale (1.....2.....3.....4.....5)

Class Overall Mean Rating = 3.63-->
Your Overall Mean Rating = 3.91-->

Item	5 Pt Score — Your Mean=x	SEM	Class Mean=c	StdDev	Raw Score — Perfect p	Yours xp/5	# of Raters
1 A. Prevntn/intervntn/whole-pt focus	3.67	.25	3.59	.308	5	3.67	6
2 B. Gathering data by history	3.83	.23	3.49	.332	5	3.83	6
3 C. Gathering data by physical exam	3.80	.24	3.55	.309	5	3.80	5
4 D. Gathering data by lab studies	3.67	.21	3.48	.229	5	3.67	6
5 E. Assessing data/arriving at DX	3.80	.27	3.53	.351	5	3.80	5
6 F. Maintain health + manage prob's	4.00	.26	3.63	.322	5	4.00	5
7 G. Interpersonal rela w H-C team	4.17	.25	3.79	.306	5	4.17	6
8 H. Work habits + personal qualities	4.00	.23	3.94	.231	5	4.00	6

Rating Scale (for a resident at this level of training)

5 = Superior performance — 1 C
4 = Very good performance — 11 C
3 = Good performance — 1 C
2 = Adequate performance — 1 C
1 = Poor performance — 11111 CCCCC

Definition of Symbols
C = class overall 5 pt score: 3.63 StdDev: .266
X = your overall 5 pt score: 3.91 SEM : .21
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 31.30 (out of perfect 40) yields: 78.3% Z= 605 Rank= 3 (out of 14). Class ave raw score 29.07

Figure 12. Individual Cumulative Score Report from a Pediatric Residency

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 5-Aug-86 14:24 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: MARCUS P. WELBY
From: Bob Glenn, MD
Re: Your test scores and cumulative scores to date

Dept: Pediatrics
Course: PL-1 Pedi Res Performance Evaluation (1985-86)

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
1	29.07	14	30-Jun-86	FACULTY RATING OF RESIDENT	1	31.30	40.00	78.25	3.91	3	605
	Currently Enrolled: 14			Total applicable weights:	1						
				Your total raw points and points possible:		31.30	40.00	———	———	———	———
				Your unweighted mean raw points, and derived rank and Z:		78.25	———	———	———	3	605
				Your weighted normalized scores, and derived rank and Z:		———	———	78.25	3.91	3	605
				The weighted mean of your Z scores is 605.0, derived rank and Z:		———	———	———	———	3	605

Residents are encouraged to come to my office to discuss their performance evaluations at any time. On a more formal basis, my office arranges for 2 or 3 appointments with me yearly for each housestaff officer for a one-to-one discussion of all of the resident's evaluations. Not only are the "rating numbers" in the resident's Individual Performance Report (e.g., the IPR in Figure 11) given to the resident but also all comments (good and bad) provided by faculty are discussed. I suggest ways the resident may go about improving his or her performance based upon the information provided in the resident's IPR. The IPR also permits the resident to see how his or her performance compares with other residents at the same level of training.

Director's Reaction and Comments

The computer prepared rating sheets and reports offer several major advantages over older manually prepared sheets. The rating sheets are rapidly prepared to our custom specifications by our Computer Services Department and quickly completed by faculty. The data on the rating sheets are easily summated, in neat, easily filed, and understandable reports. The various automatically provided reports (Figures 8-12) provide information that would be difficult and impractically time-consuming to obtain by entirely manual methods. The reports derived from the performance rating inventories are used to inform the residents, guide the improvement of their performance and also to summarize, in a quantified fashion, their skills and abilities at particular points in time. If a resident is having problems, these reports serve as documentation which can, if necessary, be used to justify reprimands or repeat rotations. If ever needed, the reports can serve as legal documents summarizing the faculty's judgement of a resident's performance.

At the end of the training program the chairman of the department must prepare an evaluation of each resident for the American Board of Pediatrics. The rating inventories and their computerized results have not only simplified this task but also have made these reports more meaningful and less subjective.

Figure 13. Rating Sheet and Inventory from a Pediatric Clerkship

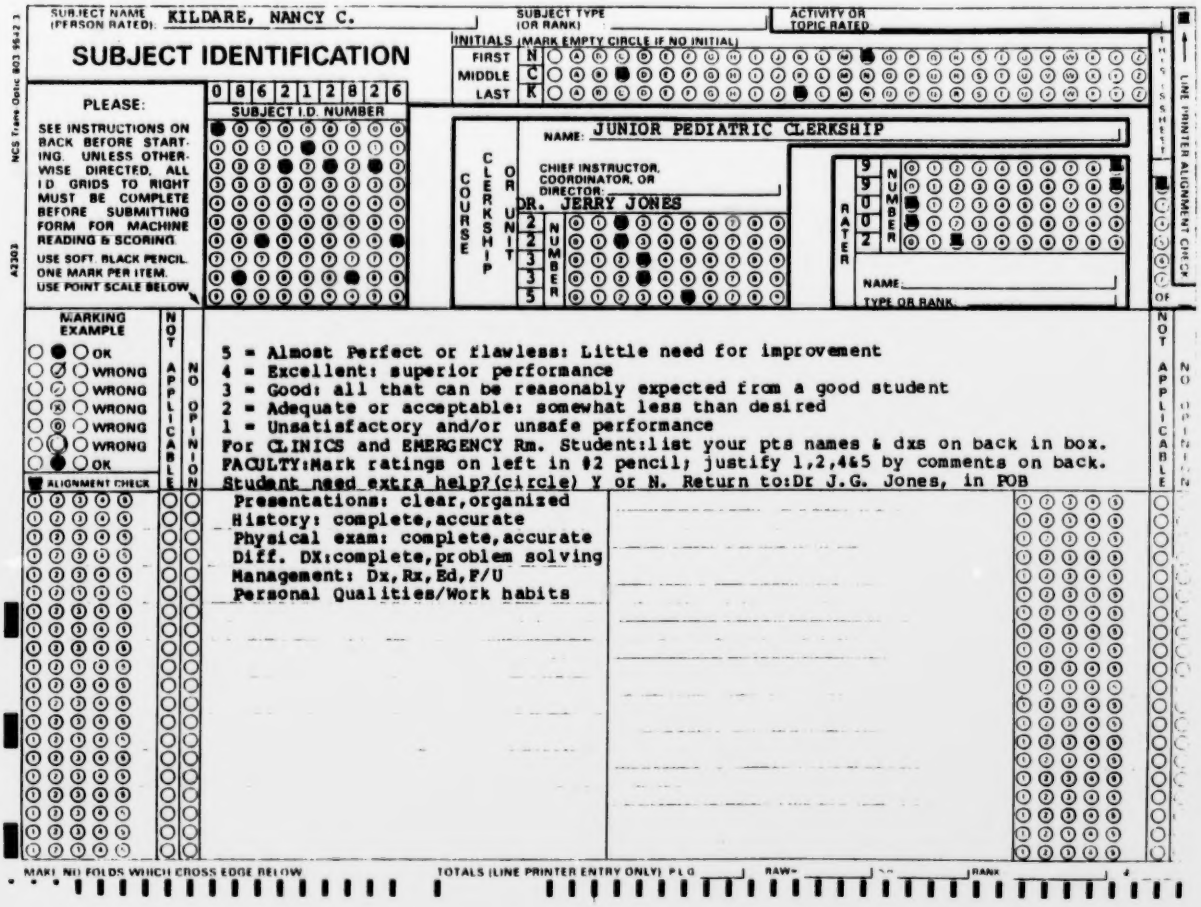

SUBJECT NAME (PERSON RATED): KILDARE, NANCY C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING. UNLESS OTHERWISE DIRECTED, ALL ID GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT, BLACK PENCIL. ONE MARK PER ITEM. USE POINT SCALE BELOW.
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
COURSE OR CLERKSHIP
NAME: JUNIOR PEDIATRIC CLERKSHIP
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR:
DR. JERRY JONES
UNIT NUMBER 2 2 3 3 5
RATER NUMBER 9 9 0 0 2
NAME:
TYPE OR RANK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
NOT APPLICABLE
NO OPINION
5 = Almost Perfect or flawless: Little need for improvement
4 = Excellent: superior performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired
1 = Unsatisfactory and/or unsafe performance
For CLINICS and EMERGENCY Rm. Student:list your pts names & dxs on back in box.
FACULTY:Mark ratings on left in #2 pencil; justify 1,2,4&5 by comments on back.
Student need extra help?(circle) Y or N. Return to:Dr J.G. Jones, in POB
ALIGNMENT CHECK
Presentations: clear,organized
History: complete,accurate
Physical exam: complete,accurate
Diff. DX:complete,problem solving
Management: Dx,Rx,Ed,F/U
Personal Qualities/Work habits
NOT APPLICABLE
NO OPINION
MARK NO FOLDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) PLG
RAW=
RANK
LINE PRINTER ALIGNMENT CHECK

Figure 14. Rating Analysis Summary from a Pediatric Clerkship

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 29-Jul-86 15:49 by the UAMS VIS/PR System (version B0) as implemented at UAMS

Test: 5 - CLINICS & EMERGENCY ROOM RATING
Instructor: DR. JERRY JONES
Course: JUNIOR PEDIATRIC CLERKSHIP

Dept: Pediatrics
Slot: 512 Phone:370-1013
Subjects rated:121 Absent: 0 Withdrawn: 14

# Rated Items & Total Points (1)		Average Score Raw	5 Pt	1 Rater Reliab.	Mean # Raters & Reliability (2)		Standard Dev Raw	5 Pt	Std Error of Measure Raw	5 Pt	Z (3)
6	6	4.3	3.55	0.10	18.0	0.67	0.3	.215	0.15	.12	57

A Flexible System For Processing Clinical Performance Ratings: Application in a Junior Year Pediatrics Clerksnip*

The Pediatric Clerkship is a required clinical rotation for all Junior medical students at UAMS. The clerkship is carried out primarily at Arkansas Children's Hospital (ACH). (ACH, the teritary pediatric care referral center for the whole state, is described in more detail in the Pediatric Residency section of this paper.) Each student also spends 5 days at UAMS's nursery. The pediatric clerkship is a rather complicated one, with more than 50 faculty physicians who teach and rate the 140 students in each Junior class. Each student rotates through three clinical areas. Groups of 22 to 25 Junior medical students spend a total of 8 weeks in pediatrics: 4 weeks on the inpatient service, 3 weeks in the clinics and emergency room, and one week in the term newborn nursery. Altogether, 15 different pediatric sub-specialty clinics are used for the clinical training of Junior year medical students. Each student is required to rotate through at least 11 of these 15 clinics.

Performance Evaluation Procedures

We have been using the computer (i.e., OTS-PR) to process both our clinical ratings and exams since 1983-84. In each clinical evaluation rendered, a 5-point scale of defined ratings is used to assess each of 6 performance criteria. The 6 criteria are appropriately different for each clinical area through which students rotate (i.e., inpatient, newborn nursery, and clinics/ER). Figure 13 gives an illustration of the inventory (printed on a rating sheet) used in the clinics and ER. Figure 14 gives an illustration of the computer analysis of the reliability of the rating process in the clinics. Figures 15 and 16-17 give examples of reports on a group of students and an individual student respectively. All the example reports (Figures 14-17) contain actual data from our clerkship during 1985-86. The data on which these examples are based included all but one group of students from the Junior class of 1985-86.

For each student, rating sheets are completed by 1 or 2 inpatient faculty attendings, 1 or 2 inpatient senior pediatric residents, a nursery faculty attending; and, outpatient faculty attendings in each of the following: a general pediatric clinic, 11 pediatric sub-specialty clinics, and the emergency room. As shown in Table 6, for the 1983-84 year about 17 rating sheets were completed on each student. Figure 14 indicates this number is near 18 for 1985-86. More importantly, the inter-rater reliability for 1985-86 (Figure 14) is slightly higher than for 1983-84 for the clinics and ER ratings.

Upon students' completion of the inpatient and nursery rotations, rating sheets are sent to the raters by the clerkship director. Students in the clinics and emergency room are responsible for giving a copy of their own rating sheet to each of their teachers/raters (these are pre-slugged with alternate student identification numbers to preserve confidentiality). Each student is provided with a supply of his or her rating sheets in convenient locations in the clinics and the emergency room. At the end of each clinic day or one-half

*Prepared by Jerry G. Jones, M.D., Clerkship Director. Address communications concerning details of this application to this author at UAMS-COM-512, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 370-1013 (Arkansas Children's Hospital).

Figure 15. Students by Rank Report from a Pediatric Clerkship

STUDENTS BY RANK (Current Rating)

Prepared 30-Jul-86 13:37 by the UAMS OIS/PR System (version B0) as implemented at UAMS Page 1

Test: 5 - CLINICS & EMERGENCY ROOM RATING
Instructor: DR. JERRY JONES
Course: JUNIOR PEDIATRIC CLERKSHIP

Dept: Pediatrics
Slot: 512 Phone:370-1013
Subjects rated:121 Absent: 0 Withdrawn: 14

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	# OF RATERS	NAME
	Averages	4.25	3.55			500	18.0	(Actual data; fictitious names)
1	283001105	5.12	4.26	1	99	833	12	RADT
2	283000529	4.89	4.07	2	98	745	22	SNFD
3	283001915	4.88	4.07	3	97	744	16	CAHJ
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
89	283001279	4.09	3.41	89	26	435	20	KILDARE, NANCY C.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
121	283001519	3.76	3.14	121	0	310	20	TSKR

Figure 16. Individual Performance Report from a Pediatric Clerkship

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 29-Jul-86 15:57 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. JERRY JONES
Re: Rating/test 5-CLINICS & EMERGENCY ROOM RATING

Dept: Pediatrics
Course: JUNIOR PEDIATRIC CLERKSHIP

Item	5 Point Scale 1.....2.....3.....4.....5	— Your 5 Pt Score — Mean=x	SEM	— Class 5 Pt Score — Mean=c	StdDev	— Raw Score — Perfect p	Yours xp/5	# of Raters
Class Overall Mean Rating = 3.55-->	C							
Your Overall Mean Rating = 3.41-->	X							
1 Presentations: clear,organized	..x	3.61	.15	3.59	.259	1	0.72	18
2 History: complete,accurate	.x	3.50	.15	3.51	.248	1	0.70	18
3 Physical exam: complete,accurate	x.c	3.29	.15	3.50	.217	1	0.66	17
4 Diff. DX:complete,problem solving	x c.	3.17	.14	3.43	.215	1	0.63	18
5 Management: Dx,Rx,Ed,F/U	x.c	3.33	.15	3.48	.231	1	0.67	15
6 Personal Qualities/Work habits	..x c	3.61	.16	3.87	.254	1	0.72	18

Rating Scale
5 = Almost Perfect or flawless: Little need for improvement
4 = Excellent: superior performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired
1 = Unsatisfactory and/or unsafe performance

Definition of Symbols
C = class overall 5 pt score: 3.55 StdDev: .215
X = your overall 5 pt score: 3.41 SEM : .12
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 4.09 (out of perfect 6) yields: 68.1% Z= 435 Rank= 89 (out of 121). Class ave raw score 4.25

Figure 17. Individual Cumulative Score Report from a Pediatric Clerkship

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 29-Jul-86 16:06 by the UAMS OIS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. JERRY JONES
Re: Your test scores and cumulative scores to date

Dept: Pediatrics
Course: JUNIOR PEDIATRIC CLERKSHIP

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
Some lines describing assessments not applicable to this subject were manually deleted.											
1	101.70	20	30-Aug-85	PEDIATRIC DEPARTMENTAL EXAM	25	104.00	138.00	75.36	3.77	9	523
2	23.67	121	01-Jul-86	INPATIENT RESIDENT RATING	10	24.00	30.00	80.00	4.00	58	510
3	22.86	121	01-Jul-86	INPATIENT FACULTY RATING	10	22.00	30.00	73.34	3.67	67	469
4	20.02	121	01-Jul-86	NURSERY RATING	10	21.00	30.00	70.00	3.50	22	551
5	4.25	121	01-Jul-86	CLINICS & EMERGENCY ROOM RATING	20	4.09	6.00	68.13	3.41	89	435
11	497.48	121	02-Jul-86	NATIONAL BOARDS	25	365.00	999.00	36.54	1.83	108	370
	Currently Enrolled:121			Total applicable weights: 100							
				Your weighted normalized scores, and derived rank and Z:		———	———	63.94	3.20	77	451
				The weighted mean of your Z scores is 463.3, derived rank and Z:		———	———	———	———	82	444

Figure 18. Facsimile of Explanatory Letter

INTERPRETING YOUR COMPUTER-GENERATED GRADE REPORTS FOR
THE PEDIATRIC CLERKSHIP

Individual Cumulative Score Report

This page provides your Z-score and rank on each of the following: departmental exam, resident and faculty attending inpatient ratings, nursery rating, and clinics/ER ratings, as well as the relative weight of each. In the lower right hand corner of the sheet is your provisional rank and Z-score for all of the five items combined. Remember that this total score is provisional, in that your Z-scores on the four clinical items will be revised at the end of the year to be based on the ratings of all the students in the junior class. Only your departmental exam Z-score will not change, since it will remain based on your rotational group alone. If you have questions about this, consult the page on grading in your clerkship orientation manual.

Prior to your National Board Examination on April 30 and May 1, 1984, I will try to provide you with a provisional Z-score based on a larger number of students' ratings. This more definitive score should be available in the Pediatric Office at UAMS on February 15, 1984.

Report for Department Examination

This page provides your Z-score on the departmental exam, your rank in your rotational group, and the average score of your class. The number of points (and questions) on which your score is based is less than 150 because questions that had poor reliability indices were omitted.

Individual Performance Reports for Clinical Activities

You should have four clinical Individual Performance Reports, one based on ratings of each of the following raters:

Resident(s) on Inpatient Service
Faculty attending(s) on Inpatient Services
Nursery attending(s)
Clinics/ER attendings

The rater(s) on which each of the four sheets is based is indicated on the fourth line on the left side of the page.

In each of these four reports, you will be interested primarily in the two columns headed "Mean Rating, 5-Point Scale". The first column, under the heading "x", shows your average rating on each of the five evaluation criteria listed on the left side of the page, numbered 2 through 6. The second column, labelled "c", provides the average rating of your entire group of students on each of these same items.

Jerry G. Jones, M.D.
Clerkship Coordinator

day, students initiate a new rating sheet by listing on its back the names and diagnoses of each patient seen, then giving the sheet to the faculty attending. The faculty person records his or her ratings of and comments about the student and places the sheet in the collection box provided in the clinical area.

Final grades in the clerkship are derived from two kinds of evaluation procedures: ratings of performance observed in various clinical settings and scores on examinations. These performance areas and their respective weights in the computation of final grades are given in Table 4.

Table 4. Components of Evaluation

Performance Ratings	Weight
Inpatient: Faculty Attending	10%
Inpatient: Senior Resident	10%
Nursery	10%
Clinics/Emergency Room	20%
Examinations	
Departmental Examination	25%
NBME Subject Exam in Pediatrics	25%

All performance ratings (i.e., the four items at top of Table 4) are cumulative through the academic year. That is, the performance ratings from each successive cohort are added to a computer file. All data in the file are processed together. Thus, as each successive cohort completes the clerkship, the report on each student that has completed the clerkship reflects data on a greater portion of the whole class. Final grades are based on the whole class' ratings, not just those in the cohort with whom a particular student trained. Students are given provisional scores on each of the performance rating areas at the end of the 8 week clerkship. A letter of explanation (Figure 18) accompanies each student's set of Individual Performance Reports (one IPR for each rating area listed in Table 4) and an Individual Cumulative Score Report (ICSR illustrated in Figure 17). Only rarely has a student requested additional explanation. The NBME subject exam in Pediatrics is given near the end of the academic year. Students are provided updated, but still provisional, reports approximately 3 months prior to that examination. A student's final score is based on a weighted average of his/her Z-scores in each rating area and on the two exams. In the ICSR (Figure 17) the student's overall score is located at the extreme lower right corner. Final computing at the end of the academic year provides final clerkship Z-scores. They are given in a Z-score Gradebook report in descending numerical order (in a report similar to Figure 15) for easy assignment of letter grades.

Director's Reaction and Comments

The ratings of the clinical performance of medical students frequently are provided subjectively by many raters with variable leniency, based on a system designed intuitively and lacking a practical means of determining its reliability. This was true of our clinical evaluations of Junior medical students on the pediatric clerkship in the academic years 1978-1979 and 1979-80. During this period, the students were graded after each clinic or emergency room session, and at the end of their inpatient service rotation. The grades were global (rather than based on ratings of individual performance criteria) and hand tabulated. Some of the results are summarized in Table 5:

Table 5. Results of Two Manual Grading Approaches

Year	Grading	Site	Final Grades
1979-80	A - F no specific definitions	Clinics/ER	20% A's, 80% B's
		Inpatient	42% A's, 59% B's
1982-83	1-5 (3= good; typical student)	Clinics/ER	2% 5's, 50% 4's, 39% 3's
		Inpatient	14% 5's, 62% 4's, 23% 3's

There were two failings of the approaches used in the two years presented in Table 5. First, general rater leniency resulted in the average of grades assigned in the whole class being higher than "C". This was incompatible with a college faculty policy that required the average course grade to be a "C". Second, evidence was convincing that the magnitude of differences in the leniency of individual raters was sometimes unacceptable. This specific problem resulted in the discard of the inpatient performance ratings from the calculation of overall clerkship grades in 1979-1980.

The negative experiences we had with evaluation of clinical performance (using a manual system) in 1979-80 resulted in our basing the grades for the academic years 1980-1981 and 1981-1982 primarily on scores of the students' two examinations. Both of these examinations had contributed to the students' final clerkship grades in the previous two years. The clinical performance of the students deteriorated severely when it weighted relatively lightly in the overall clerkship score. The students spent more time reading and less time with patients because they recognized this to be more helpful in obtaining a good grade. The students' evaluations of the course also indicated that they recognized this was reducing the quality of the clinical learning that occurred. Figure 14 provides an example of a rating analysis report from current (1985-86) data. Such reports on the ratings beginning in 1983-84 indicated that the reliability of the clinical ratings was high enough to justify our weighting them more heavily in the computation of our overall clerkship grade. Our overall grading policy and its justification is much the same as that reported by Littlefield and his colleagues (1981). Table 6 summarizes the results we achieved in 1983-84.

Table 6. Reliability of Evaluation Components in 1983-84.

Performance Component	Average rating sheets per student	Reliability
Clinics/ER	17.4	0.58
All Clinical Experience	21.8	0.62
Departmental Exam	NA	0.77
External Exam	NA	0.85
Weighted Total	NA	0.90

The advantages of the current system of evaluation of clinical performance are as follows:

1. Six specific criteria are to be rated each time clinical performance is assessed. This was a practical impossibility for us prior to 1983-84 in the

manually tabulated system. Use of multiple criteria assist the faculty in providing a more complete assessment of students' performance. The summary of these more detailed assessments on the computer generated IPR reports (which are given to the students at the end of the clerkship) permit the students to assess their relative strengths and weaknesses in clinical performance against both the stipulated standard given on the rating sheet (Figure 13) and IPR (Figure 16) as well as the average performance of their peers.

2. The average clinical performance scores of the students now form an approximately bell-shaped curve, facilitating assigning a final average class grade of "C", as required by our College of Medicine.

3. The reliability of the clinic performance scores is quantified.

4. The result of future changes in the method of assessment of clinical performance can be evaluated objectively.

Some rater "grade inflation" has occurred since we began using OTS-PR to process ratings in 1983. Specific instances of extreme rater leniency have been identified, particularly in the inpatient area. Each criterion of the inpatient service rating inventory was evaluated, and none seemed sufficiently contributory to rater grade inflation to justify alteration of the current inventory. As a first step towards correcting the excessive rater leniency problem, we have provided to each faculty rater a profile of his or her student inpatient ratings in each of the last 3 academic years compared with the overall average student ratings for those years. The results reported by Littlefield and his associates (1981) give little basis to hope that feedback to raters alone will correct the leniency problem. If significant rater grade inflation or excessive rater leniency is identified at the end of the current academic year, additional measures (such as "handicapping" ratings a la Delk, Cason, and Reese, 1985) may be taken.

The advantages seem to justify the time required to administer the current rating system. However, I could recommend it to the director of a clerkship similar to ours only if he or she has good computing and educational professional support, and a very well-organized and highly reliable secretary.

I suspect our clerkship's performance rating reports would be helpful in selection of residents in pediatrics, medicine, and family practice programs. However, we have not yet formally evaluated this suspicion.

A Flexible System For Processing Clinical Performance Ratings: Application in a Junior Year Specialties Clerkship*

The Junior Specialties Clerkship is a required Junior medical student rotation. Each student spends four weeks in either Orthopedics or Otorhinolaryngology and another four weeks in either Neurology or Neurosurgery. Every eight weeks, 22 students are evenly divided among four different combinations of clinical rotations. The clinical experience is consequently marked by a wide diversity of settings ranging from pediatric outpatient clinics to adult ward services and operating rooms. What makes the Specialties Clerkship unique among the Junior rotations is this variation in clinical exposure and the large number of faculty and residents responsible for teaching and evaluation.

Performance Evaluation Procedures

Students' grades in the course are assigned on the basis of a final written examination covering didactic material assigned to the entire group (60% of grade) and on the basis of performance rating scores in their two clinical rotations (20% each). The overall course score is computed as a weighted average of Z-scores earned on the exam and in the two clinical rotations. The performance rating scores from Neurology are submitted as a single numerical grade by that Department. The remaining three clinical departments use the rating sheets processed by the Performance Rating System. Residents and faculty members are asked to complete computer generated rating sheets on each student they taught. Each department selects the areas to be evaluated so that the system is customized to fit their needs. Figure 19 illustrates the evaluation inventory used in Neurosurgery. In general, two to five raters in each department submit sheets on each student based on the raters' direct observations. No special information about completing the rating sheets is provided to the rater.

The results of the ratings are supplied to the students by posting a report in which students are designated by confidential identification numbers and performance is expressed in Z-scores and rank in class. Students who have questions concerning their scores meet with the course director to review their Individual Performance Reports (Figure 22) and Individual Cumulative Score Reports (Figure 23). If they still have questions, they can review the original rating sheets. In addition, written comments by raters are solicited and these are available to the student for review on request. We retain the rating sheets, IPRs, and ICSRs in individual student files and the class gradebook reports in our permanent files. The Dean's office receives only the final course grade and copies of all raters' written comments.

Director's Reactions and Comments

In a multi-disciplinary course such as Specialties, the OTS-PR System allows multiple, diverse assessments to be quickly and fairly combined to yield clinical performance scores. The clinical scores can be easily combined with exam scores to produce an overall score from which final grades may be assigned.

*Prepared by Kimberly A. Golden, M.D. Clerkship Director. Address communications concerning details of this application to this author at UAMS-COM-515, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 661-6114.

Figure 19. Rating Sheet and Inventory from a Specialties Clerkship

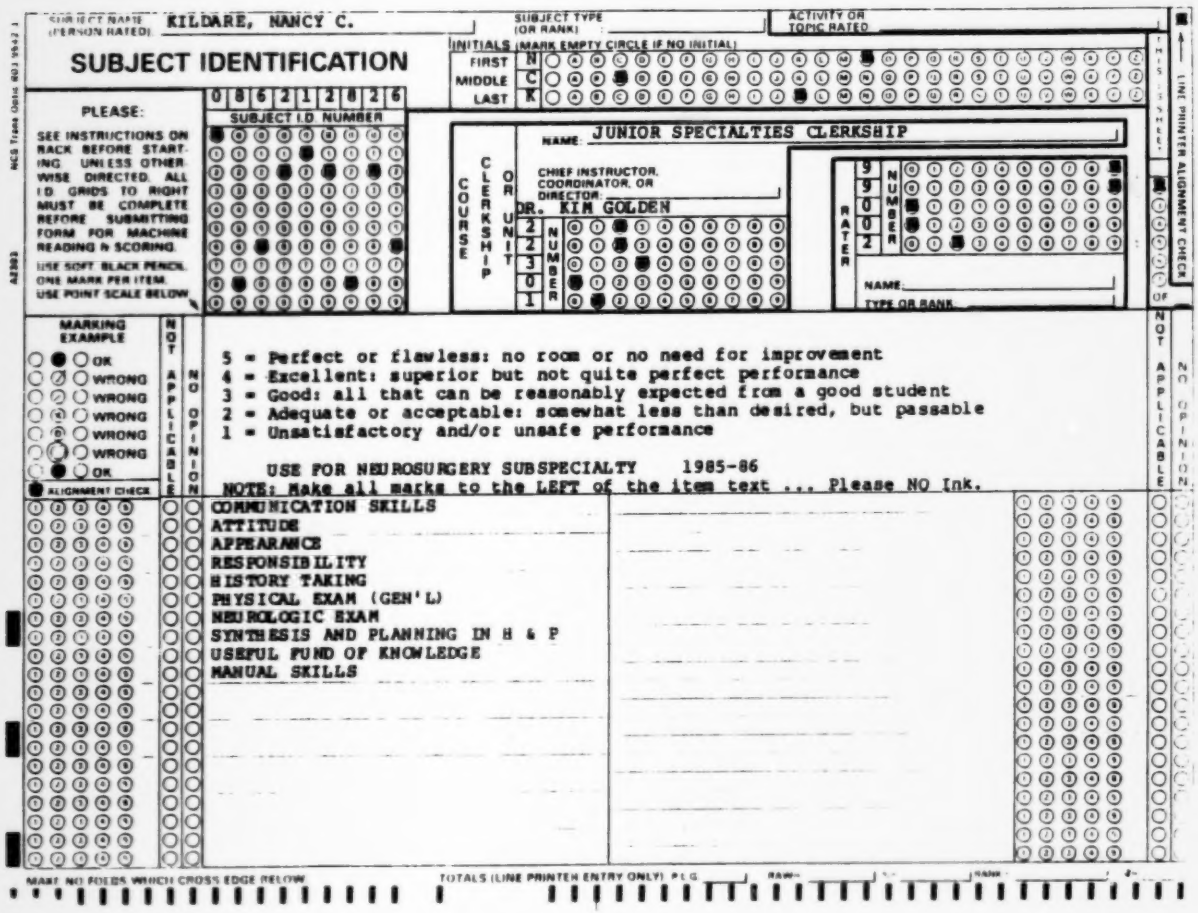

SUBJECT NAME (PERSON RATED) KILDARE, NANCY C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING UNLESS OTHERWISE DIRECTED. ALL ID GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT BLACK PENCIL ONE MARK PER ITEM. USE POINT SCALE BELOW.
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
COURSE OR UNIT CLERKSHIP NUMBER 2 2 3 0 1
NAME: JUNIOR SPECIALTIES CLERKSHIP
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR:
DR. KIM GOLDEN
RATER NUMBER 9 9 0 0 2
NAME:
TYPE OR RANK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsafe performance
USE FOR NEUROSURGERY SUBSPECIALTY 1985-86
NOTE: Make all marks to the LEFT of the item text ... Please NO Ink.
COMMUNICATION SKILLS
ATTITUDE
APPEARANCE
RESPONSIBILITY
HISTORY TAKING
PHYSICAL EXAM (GEN'L)
NEUROLOGIC EXAM
SYNTHESIS AND PLANNING IN H & P
USEFUL FUND OF KNOWLEDGE
MANUAL SKILLS
MARK NO FOLDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) P.I.G.

Figure 20. Rating Analysis Summary from a Specialties Clerkship

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 29-Jul-86 14:40 by the UAMS UIS/PR System (version 80) as implemented at UAMS

Test: 8 - NEUROSURGERY RATING
Instructor: DR. KIM GOLDEN
Course: JUNIOR SPECIALTIES CLERKSHIP

Dept: Anesthesiology
Slot: 515 Phone:661-6114
Subjects rated: 60 Absent: 68 Withdrawn: 8

Category	# Rated Items & Total Points (1)		Average Score Raw	Average Score 5 Pt	1 Rater Reliab.	Mean # Raters & Reliability (2)		Standard Dev Raw	Standard Dev 5 Pt	Std Error of Measure Raw	Std Error of Measure 5 Pt	Measure Z (3)
1 COMMUNICATION SKILLS	1	5	3.8	3.80	0.21	3.7	0.49	0.5	.513	0.37	.37	71
2 ATTITUDE	1	5	4.1	4.08	0.20	3.7	0.48	0.5	.455	0.33	.33	72
3 APPEARANCE	1	5	4.0	3.97	0.17	3.7	0.43	0.4	.428	0.32	.32	75
4 RESPONSIBILITY	1	5	3.9	3.92	0.21	3.5	0.48	0.5	.540	0.39	.39	72
5 HISTORY TAKING	1	5	3.6	3.64	0.35	2.8	0.60	0.6	.555	0.35	.35	63
6 PHYSICAL EXAM (GEN'L)	1	5	3.7	3.67	0.36	2.7	0.60	0.6	.552	0.35	.35	63
7 NEUROLOGIC EXAM	1	5	3.6	3.57	0.24	3.0	0.49	0.5	.494	0.35	.35	72
8 SYNTHESIS AND PLANNING IN H & P	1	5	3.6	3.63	0.18	3.2	0.41	0.5	.481	0.37	.37	77
9 USEFUL FUND OF KNOWLEDGE	1	5	3.6	3.61	0.21	3.6	0.49	0.4	.435	0.31	.31	72
10 MANUAL SKILLS	1	5	3.6	3.64	0.12	2.1	0.22	0.5	.486	0.43	.43	88
Overall	10	50	37.4	3.74	0.30	3.8	0.61	4.0	.398	2.47	.25	62

Figure 21. Students by Rank Report from a Specialties Clerkship

STUDENTS BY RANK (Current Rating)

Prepared 30-Jul-86 14:27 by the UAMS OTS/PR System (version B0) as implemented at UAMS Page 1

Test: 8 - NEUROSURGERY RATING
Instructor: DR. KIM GOLDEN
Course: JUNIOR SPECIALTIES CLERKSHIP

Dept: Anesthesiology
Slot: 515 Phone:661-6114
Subjects rated: 60 Absent: 68 Withdrawn: 8

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	# OF RATERS	NAME
	Averages	37.41	3.74			500	3.8	(Actual data; fictitious names)
1	283000719	49.33	4.93	1	98	799	3	ASDD
2	283001980	46.22	4.62	2	96	721	3	BEFP
3	283001915	45.92	4.59	3	95	714	5	CNLJ
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
18	283001279	39.17	3.92	18	70	544	4	KILDARE, NANCY C.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
60	283000693	28.67	2.87	60	0	281	3	ADHF

Figure 22. Individual Performance Report from a Specialties Clerkship

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 29-Jul-86 14:53 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. KIM GOLDEN
Re: Rating/test 8-NEUROSURGERY RATING

Dept: Anesthesiology
Course: JUNIOR SPECIALTIES CLERKSHIP

Item	5 Point Scale 1.....2.....3.....4.....5	5 Pt Score — Your Mean=x	SEM	Class Mean=c	Class StdDev	Raw Score Perfect p	Raw Score Yours xp/5	# of Raters
Class Overall Mean Rating = 3.74	C							
Your Overall Mean Rating = 3.92	X							
1 COMMUNICATION SKILLS	.x.	3.75	.36	3.80	.513	5	3.75	4
2 ATTITUDE	.-.C—x—	4.75	.32	4.08	.455	5	4.75	4
3 APPEARANCE	.-.x	4.00	.32	3.97	.428	5	4.00	4
4 RESPONSIBILITY	.C—x—	4.50	.38	3.92	.540	5	4.50	4
5 HISTORY TAKING	xc.-.	3.50	.39	3.64	.555	5	3.50	2
6 PHYSICAL EXAM (GEN'L)	x—C-.	3.00	.38	3.67	.552	5	3.00	2
7 NEUROLOGIC EXAM	x—C.-.	3.00	.39	3.57	.494	5	3.00	2
8 SYNTHESIS AND PLANNING IN H & P	x—C.-.	3.33	.37	3.63	.481	5	3.33	3
9 USEFUL FUND OF KNOWLEDGE	xc.-.	3.50	.30	3.61	.435	5	3.50	4
10 MANUAL SKILLS	cx-.	3.67	.41	3.64	.486	5	3.67	3

Rating Scale
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsure performance

Definition of Symbols
C = class overall 5 pt score: 3.74 StdDev: .398
X = your overall 5 pt score: 3.92 SEM : .24
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 39.17 (out of perfect 50) yields: 78.3% Z= 544 Rank= 18 (out of 60). Class ave raw score 37.41

Figure 23. Individual Cumulative Score Report from a Specialties Clerkship

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 29-Jul-86 15:03 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. KIM GOLDEN
Re: Your test scores and cumulative scores to date

Dept: Anesthesiology
Course: JUNIOR SPECIALTIES CLERKSHIP

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
Some lines describing assessments not applicable to this subject were manually deleted.											
5	8.12	60	02-Jul-86	NEUROLOGY RATING	1	ABSENT	12.00				
6	38.72	58	02-Jul-86	ORTHOPEDICS RATING	1	44.50	45.00	98.89	4.94	3	624
7	34.84	61	02-Jul-86	ENT RATING	1	ABSENT	45.00				
8	37.41	60	02-Jul-86	NEUROSURGERY RATING	1	39.17	50.00	78.33	3.92	18	544
9	503.03	121	02-Jul-86	MULTI-DEPT. EXAM	3	305.00	999.00	30.53	1.53	116	298
	Currently Enrolled:128			Total applicable weights:	5						
				Your weighted normalized scores, and derived rank and Z:		———	———	53.76	2.69	105	403
				The weighted mean of your Z scores is 412.4, derived rank and Z:						110	374

With an entirely manual system, evaluation and processing would require considerably more time on three levels: faculty/rater, departmental coordinator, and Course Director. In addition to requiring much more time, grading conferences are subject to group bias which can prevent fair, independent evaluation of students. My impression as Course Director after using the Individual Performance Reports in meetings with students is that instructors tend to be more honest in their computer ratings than they can be in face-to-face confrontation with students or even in their written comments about student performance. Another advantage to the IPR's is the ease of use; they provide an uncluttered graphic representation to the student of his performance relative to his classmates. This particular attribute would be quite helpful to me in the evaluation of residency candidates, because performance relative to peers is frequently difficult to evaluate with transcripts and Dean's letters alone. Another advantage to the system in a multi-departmental course like ours is that each department is free to choose to use or not use the system; the fact that one department chooses not to use the machine scannable rating sheets does not preclude another from doing so.

The OTS-PR System provides an easy-to-understand and easy-to-use means of evaluating students and providing that information to them. Grading within a department is still unfortunately subject to variability in rater stringency so that especially if few raters are used, a student's grade can be distorted. On the other hand, using Z-scores computed from the ratings provided by each department, grades are corrected for differences in difficulty (i.e., mean) and variability (i.e., standard deviation) of ratings from department to department. In this way students who select a rotation known for tougher grading, for example, are not penalized; nor, are those who chose an easy grader given a free ride.

A Flexible System For Processing Clinical Performance Ratings: Application in a Junior Year Surgery Clerkship*

The Junior Year Surgery Clerkship is based at three Little Rock nospitals: University Hospital (UH), Arkansas Childrens' Hospital (ACH), John L. McClellan Veterans' Administration Hospital (VAH). All students spend time in both inpatient and outpatient contact. The Surgery Clerkship is 8 weeks long. A student spends 4 weeks at UH; and, 4 weeks at either VAH or ACH. Six cohorts of about 24 students each rotate through the clerkship each year. Each cohort is taught, observed and evaluated by 6 different groups of staff and housestaff: one group at ACH, two at UH, and three at VAH.

Performance Evaluation Procedures

The formal evaluation of each student in the Surgery Clerkship is based on the scores from two subjective evaluations (i.e., rating sheets illustrated in Figure 24) and three multiple-choice objective exams. One rating sheet is completed at the end of each of the two four-week blocks of clinical instruction in the clerkship. The two rating sheets are completed by the individuals having the most student contact time, usually the chief resident or faculty attending. Information obtained in the subjective evaluations (i.e., ratings given and comments made) is available to the student approximately 7 to 10 days after completion of a four week rotation.

The two subjective evaluations account for one-half of the student's final clerkship grade. The rating inventory asks that the rater evaluate the performance of the student in performance categories: knowledge, interpersonal skills, and carrying out assigned tasks. The computer system allows the rater either to provide a rating for each of the 5 to 10 items (criteria) in each category or to assign only an overall score for each category. However, in the example provided in Figure 27, all individual criteria have been weighted zero so the rater can record ratings only on the "overall" items for each category. If the rater records the mark on the wrong line, as illustrated in Figure 27, line 14, the effect is as if no response were made in that category by that rater. (This error did not occur in the actual data. It is included for illustrative purposes only.)

The three objective evaluations consist of two locally written exams (Cancer and Surgery Clerkship Final) and the Surgery subject exam from the National Board of Medical Examiners. In computing a student's overall score in the clerkship these grades are weighted, respectively, 5%, 20%, and 25%. The Cancer Exam is completed at the student's convenience, but sometime during the clerkship. The Final exam is given to each cohort of students at the end of their eight week rotation through the clerkship. The NBME exam is given to all students at the end of the Junior year. At the end of the year, students are ranked based on their cumulative weighted average percent scores assigned on the two clinical ratings and the three exams. Course grades are assigned according to college faculty policy, such that the class average grade is "C" (i.e., class average falls between 1.5 and 2.49 out of 4).

*Prepared by Nicholas P. Lang, M.D. Clerkship Director. Address communications concerning details of this application to this author at UAMS-COM-520, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 661-6186.

Figure 24. Rating Sheet and Inventory from a Surgery Clerkship

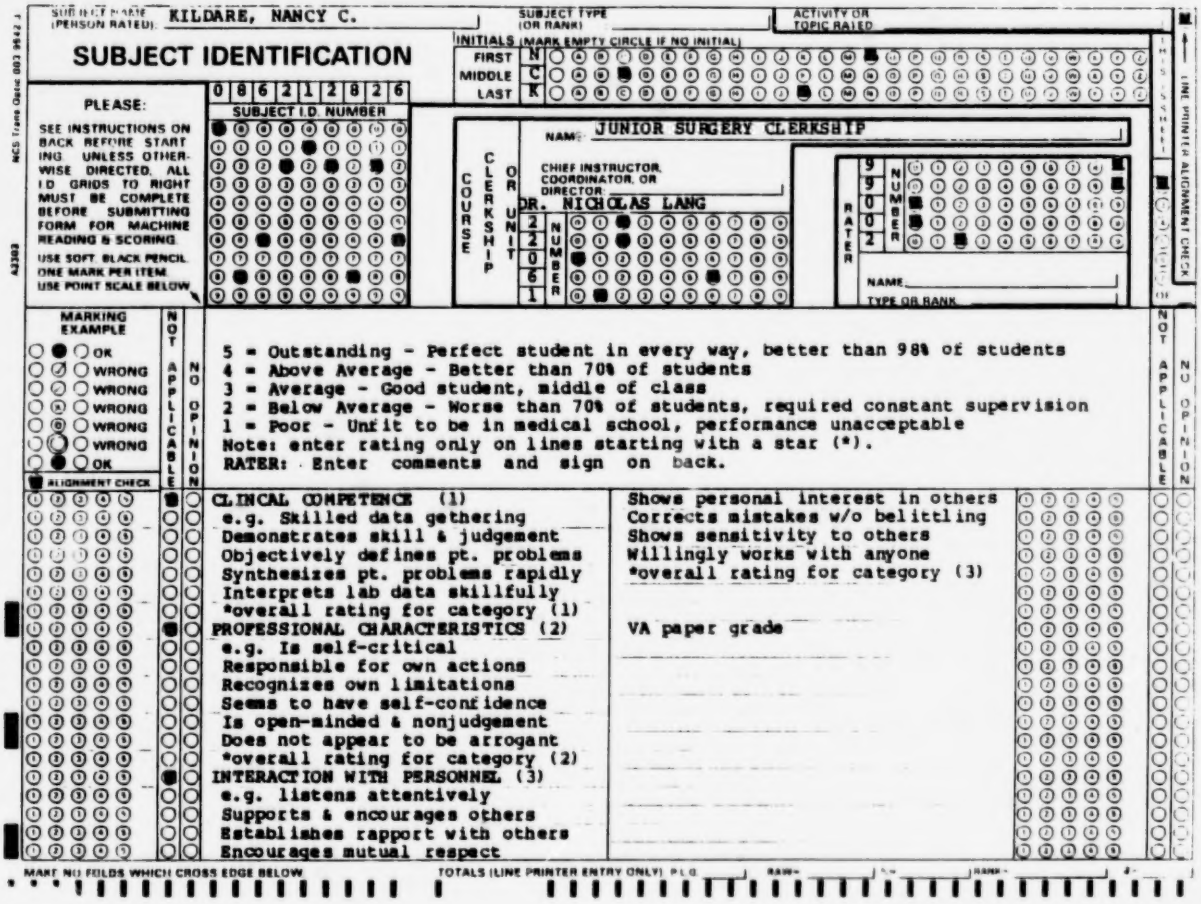

SUBJECT NAME (PERSON RATED): KILDARE, NANCY C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING. UNLESS OTHERWISE DIRECTED, ALL I.D. GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT BLACK PENCIL. ONE MARK PER ITEM.
USE POINT SCALE BELOW
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
COURSE
CLERKSHIP OR UNIT
NAME: JUNIOR SURGERY CLERKSHIP
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR
DR. NICHOLAS LANG
2 2 0 6 1
NUMBER
RATER
9 9 0 0 2
NUMBER
NAME
TYPE OR RANK
LINE PRINTER ALIGNMENT CHECK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Outstanding - Perfect student in every way, better than 98% of students
4 = Above Average - Better than 70% of students
3 = Average - Good student, middle of class
2 = Below Average - Worse than 70% of students, required constant supervision
1 = Poor - Unfit to be in medical school, performance unacceptable
Note: enter rating only on lines starting with a star (*).
RATER: Enter comments and sign on back.
CLINICAL COMPETENCE (1)
e.g. Skilled data gethering
Demonstrates skill & judgement
Objectively defines pt. problems
Synthesizes pt. problems rapidly
Interprets lab data skillfully
*overall rating for category (1)
PROFESSIONAL CHARACTERISTICS (2)
e.g. Is self-critical
Responsible for own actions
Recognizes own limitations
Seems to have self-confidence
Is open-minded & nonjudgement
Does not appear to be arrogant
*overall rating for category (2)
INTERACTION WITH PERSONNEL (3)
e.g. listens attentively
Supports & encourages others
Establishes rapport with others
Encourages mutual respect
Shows personal interest in others
Corrects mistakes w/o belittling
Shows sensitivity to others
Willingly works with anyone
*overall rating for category (3)
VA paper grade
MARK NO FIELDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) P.L.O. RAW= %= RANK=

Figure 25. Rating Analysis Summary from a Surgery Clerkship

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 5-Aug-86 14:57 by the UAMS VIS/PR System (version 80) as implemented at UAMS

Test: 13 - VAH RATING
Instructor: DR. NICHOLAS LANG
Course: JUNIOR SURGERY CLERKSHIP

Dept: Surgery
Slot: 520 Phone:661-6186
Subjects rated: 20 Absent:114 Withdrawn: 4

	Category	# Rated Items & Total Points (1)		Average Score Raw	5 Pt	1 Rater Reliab.	Mean # Raters & Reliability (2)		Standard Dev Raw	5 Pt	Std Error of Measure Raw	5 Pt	% (3)
1	CLINICAL COMPETENCE (1)	6	5	3.6	3.60	0.00	1.0	N/A	0.8	.821	0.82	.82	100
2	PROFESSIONAL CHARACTERISTICS (2)	7	5	3.7	3.72	0.00	1.0	N/A	0.8	.752	0.75	.75	100
3	INTERACTION WITH PERSONNEL (3)	12	10	8.5	4.25	0.00	1.0	N/A	1.1	.574	1.15	.57	100
	Overall	25	20	15.8	3.95	0.00	1.0	0.00	2.4	.596	2.38	.60	100

Figure 26. Students by Rank Report from a Surgery Clerkship

STUDENTS BY RANK (Current Rating) Prepared 5-Aug-86 14:58 by the UAMS OIS/PR System (version 80) as implemented at UAMS Page 1

Test: 13 - VAH RATING
Instructor: DR. NICHOLAS LANG
Course: JUNIOR SURGERY CLERKSHIP

Dept: Surgery
Slot: 520 Phone:661-6186
Subjects rated: 20 Absent:114 Withdrawn: 4

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	# OF RATERS	NAME
	Averages	15.82	3.95			500	1.0	(Actual data; fictitious names)
1	283001980	20.00	5.00	1	95	676	1	BEFP
2	283001741	19.00	4.75	2	85	634	1	YJTP
3	283001022	19.00	4.75	2	85	634	1	ZSWJ
4	283001279	18.67	4.67	4	80	620	1	KILDARE, NANCY C.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
20	283001717	10.67	2.67	20	0	284	1	AKHK

Figure 27. Individual Performance Report from a Surgery Clerkship

INDIVIDUAL PERFORMANCE REPORT (Current Rating) Prepared 5-Aug-86 14:59 by the UAMS OIS/PR System (version 80) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. NICHOLAS LANG
Re: Rating/test 13-VAH RATING

Dept: Surgery
Course: JUNIOR SURGERY CLERKSHIP

	Item	5 Point Scale 1.....2.....3.....4.....5	5 Pt Score — Your Mean=x	SEM	Class Mean=c	StdDev	Raw Score Perfect p	Yours xp/5	# of Raters
	Class Overall Mean Rating = 3.95 ⟹	C							
	Your Overall Mean Rating = 4.67 ⟹	X							
1	CLINICAL COMPETENCE (1)		(4.00)	(.82)	(3.60)	(.821)	(5)	(4.00)	(1)
2	e.g. Skilled data gathering		N/A	N/A	N/A	N/A	0	N/A	0
3	Demonstrates skill & judgement		N/A	N/A	N/A	N/A	0	N/A	0
4	Objectively defines pt. problems		N/A	N/A	N/A	N/A	0	N/A	0
5	Synthesizes pt. problems rapidly		N/A	N/A	N/A	N/A	0	N/A	0
6	Interprets lab data skillfully		N/A	N/A	N/A	N/A	0	N/A	0
7	*overall rating for category (1)	c x	4.00	.82	3.60	.821	5	4.00	1
8	PROFESSIONAL CHARACTERISTICS (2)		(N/A)	(N/A)	(3.72)	(.752)	(5)	(N/A)	(0)
9	e.g. Is self-critical		N/A	N/A	N/A	N/A	0	N/A	0
10	Responsible for own actions		N/A	N/A	N/A	N/A	0	N/A	0
11	Recognizes own limitations		N/A	N/A	N/A	N/A	0	N/A	0
12	Seems to have self-confidence		N/A	N/A	N/A	N/A	0	N/A	0
13	Is open-minded & nonjudgement		N/A	N/A	N/A	N/A	0	N/A	0
14	Does not appear to be arrogant		4.00	N/A	.00	N/A	0	0.00	1
15	*overall rating for category (2)	c	N/A	N/A	3.72	.752	5	N/A	0
16	INTERACTION WITH PERSONNEL (3)		(5.00)	(.57)	(4.15)	(.574)	(10)	(10.00)	(1)
17	e.g. Listens attentively		N/A	N/A	N/A	N/A	0	N/A	0
18	Supports & encourages others		N/A	N/A	N/A	N/A	0	N/A	0
19	Establishes rapport with others		N/A	N/A	N/A	N/A	0	N/A	0
20	Encourages mutual respect		N/A	N/A	N/A	N/A	0	N/A	0
21	Shows personal interest in others		N/A	N/A	N/A	N/A	0	N/A	0
22	Corrects mistakes w/o belittling		N/A	N/A	N/A	N/A	0	N/A	0
23	Shows sensitivity to others		N/A	N/A	N/A	N/A	0	N/A	0
24	Willingly works with anyone		N/A	N/A	N/A	N/A	0	N/A	0
25	*overall rating for category (3)	c x	5.00	.86	3.70	.865	5	5.00	1
26			N/A	N/A	N/A	N/A	0	N/A	0
27			N/A	N/A	N/A	N/A	0	N/A	0
28	VA paper grade	c x	5.00	.52	4.80	.523	5	5.00	1

Rating Scale
5 = Outstanding - Perfect student in every way, better than 98% of students
4 = Above Average - Better than 70% of students
3 = Average - Good student, middle of class
2 = Below Average - Worse than 70% of students, required constant supervision
1 = Poor - Unfit to be in medical school, performance unacceptable

Definition of Symbols
C = class overall 5 pt score: 3.95 StdDev: .596
X = your overall 5 pt score: 4.67 SEM : .60
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 18.67 (out of perfect 20) yields: 93.3% T= 620 Rank= 4 (out of 20). Class ave raw score 15.82

Figure 28. Individual Cumulative Score Report from a Surgery Clerkship

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 5-Aug-86 15:02 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. NICHOLAS LANG
Re: Your test scores and cumulative scores to date

Dept: Surgery
Course: JUNIOR SURGERY CLERKSHIP

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
Some lines describing assessments not applicable to this subject were manually deleted.											
11	27.35	20	25-Oct-85	SURGERY EXAM	20	31.00	50.00	62.00	3.10	3	599
12	37.65	20	25-Oct-85	CANCER EXAM	5	39.00	50.00	78.00	3.90	6	542
13	15.82	20	02-Sep-85	VAH RATING	25	18.67	20.00	93.34	4.67	4	620
14	14.38	20	02-Sep-85	ACH/UAMS RATING	45	18.67	20.00	93.33	4.67	1	655
28	31.89	19	02-May-86	SURGERY FINAL EXAM	20	ABSENT	50.00				
29	N/A	0	05-May-86	CANCER PRACTICE EXAM	0	N/A	N/A				
30	450.21	117	16-Jun-86	NATIONAL BOARDS	25	315.00	999.00	31.53	1.58	103	373
	Currently Enrolled:134			Total applicable weights: 100							
				Your weighted normalized scores, and derived rank and Z:		———	———	70.85	3.54	27	574
				The weighted mean of your Z scores is 558.9, derived rank and Z:		———	———	———	———	21	595

Director's Reaction and Comments

The OTS-PR system has provided us with a method for more uniformly evaluating students at multiple hospital sites. The OTS-PR system gives the clerkship director great flexibility in student evaluation. Additional subjective evaluations, for example, can be easily included with minimal effort.

The first report of subjective evaluations, at the end of the first 4 weeks, serves to inform students of poor performance with sufficient time for them to take corrective action. For all but the last group of students, the objective score at the end of the clerkship (on the Surgery final exam) gives the student some information about his or her knowledge base in time to study and correct deficiencies before taking the NBME subject examination at the end of the Junior year.

Since the data are entered on optically read forms and all calculations are performed automatically, a major time saving occurs. The reports used to inform students of their performance include the posting report which is easy to use to communicate results to all students (similar to Figure 26 but without student names); and, individual reports (Figure 27) which provide the student detailed comparative information on his cohort's average performance.

A Flexible System For Processing Clinical Performance Ratings: Application in a Junior Year Obstetrics/Gynecology Clerksnip*

The Clerkship in Obstetrics and Gynecology is a six week period during which 17 to 18 students rotate through 3 weeks of Obstetrics and 1 week each of Gynecology, Gyn Tumor, and Ambulatory Clinics. Students attend lectures from 3 to 5 P.M. each day and have clinical responsibilities on the ward or in the delivery suite during the remainder of the day. Students are on night call every third night. In addition to formal lectures students are expected to work up patients, present small reports, write progress notes, attend grand rounds, and participate in small group conferences.

Performance Evaluation Procedures

At the end of each section of the clerksnip, the residents and staff physician(s) who worked with the students evaluate the students' clinical performance on the service using a rating sheet like that illustrated in Figure 29. At the end of the six weeks rotation the students are given a departmental final examination. The questions for the examinations are drawn from a data base of over 700 questions. Each examination is scored by computer and a Z-score assigned. This Z-score is used in the determination of the final ranking since each group takes a different departmental final examination. All examinations and evaluation rating sheets are computer scored. At the end of the six week rotation each student receives a report (Figure 33) that includes his Z-score in the departmental final exam, a self explanatory summary of his ratings on each clinical evaluation (Figure 32), and a copy of the subjective comments made by residents and attendings on each service. Since the rankings will change after each six weeks rotation as more students are ranked, students are not given interim rankings. The National Board subject exam in OB/GYN is given to Junior students as non-candidates at the end of the year. In the computation of the clerkship grade, the scores are weighted as shown in Table 7.

Table 7. Components of Evaluation

Examinations	Weight
Departmental Exam	25
National Board	25
Performance Ratings	
Obstetrical clinical evaluation	25
Gyn clinical evaluation	8
Gyn Tumor clinical evaluation	8
Clinics clinical evaluation	8

After all performance ratings and test scores are in and the weightings taken into account, the students are ranked and final grades are assigned. The distribution of letter grades will vary each year, but by college policy, the average cannot be higher than 2.5/4.0.

*Prepared by K. Lee Lee Doyle, Ph.D. Clerksnip Director. Address communications concerning details of this application to this author at UAMS-COM-518, 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 661-5753.

Figure 29. Rating Sheet and Inventory from an Obstetrics/Gynecology Clerkship

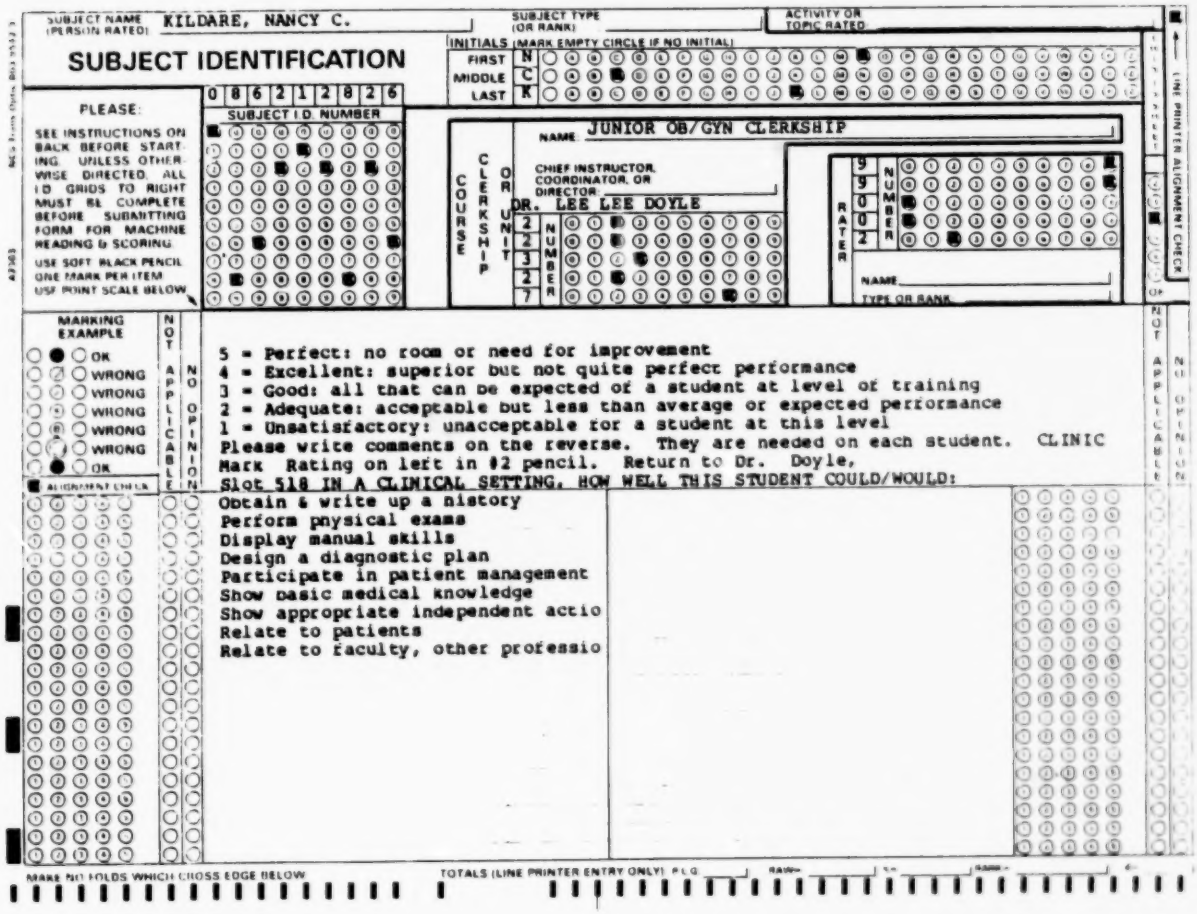

SUBJECT NAME (PERSON RATED) KILDARE, NANCY C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
SUBJECT IDENTIFICATION
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
PLEASE:
SEE INSTRUCTIONS ON BACK BEFORE STARTING UNLESS OTHERWISE DIRECTED. ALL I.D. GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING.
USE SOFT BLACK PENCIL ONE MARK PER ITEM
USE POINT SCALE BELOW
COURSE
NAME JUNIOR OB/GYN CLERKSHIP
CLERKSHIP UNIT NUMBER 2 2 3 2 7
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR
DR. LEE LEE DOYLE
RATER NUMBER 9 9 0 0 2
NAME
TYPE OR RANK
LINE PRINTER ALIGNMENT CHECK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Perfect: no room or need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be expected of a student at level of training
2 = Adequate: acceptable but less than average or expected performance
1 = Unsatisfactory: unacceptable for a student at this level
Please write comments on the reverse. They are needed on each student. CLINIC
Mark Rating on left in #2 pencil. Return to Dr. Doyle,
Slot 518 IN A CLINICAL SETTING, HOW WELL THIS STUDENT COULD/WOULD:
Obtain & write up a history
Perform physical exams
Display manual skills
Design a diagnostic plan
Participate in patient management
Show basic medical knowledge
Show appropriate independent actio
Relate to patients
Relate to faculty, other professio
MARK ALL FOLDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) P.I.Q. RAW

Figure 30. Rating Analysis Summary from an Obstetrics/Gynecology Clerkship

RATING ANALYSIS SUMMARY (Current Rating)

Prepared 29-Jul-86 14:18 by the UAMS CTS/PR System (version 80) as implemented at UAMS

Test: 5 - CLINIC RATING
Instructor: DR. LEE LEE DOYLE
Course: JUNIOR OB/GYN CLERKSHIP

Dept: Ob / Gyn
Sloc: 518 Phone:661-5753
Subjects rated:125 Absent: 5 Withdrawn: 7

# Rated Items & Total Points (I)		Average Score Raw	5 Pt	1 Rater Reliab.	Mean # Raters & Reliability (2)		Standard Dev Raw	5 Pt	Std Error of Measure Raw	5 Pt	% (3)
9	9	6.4	3.55	0.00	1.0	0.00	0.9	.499	0.90	.50	100

Figure 31. Students by Rank Report from an Obstetrics/Gynecology Clerkship

STUDENTS BY RANK (Current Rating)

Prepared 30-Jul-86 14:06 by the UAMS OTS/PR System (version B0) as implemented at UAMS Page 1

Test: 5 - CLINIC RATING
Instructor: DR. LEE LEE DOYLE
Course: JUNIOR OB/GYN CLERKSHIP

Dept: Ob / Gyn
Slot: 518 Phone:661-5753
Subjects rated:125 Absent: 5 Withdrawn: 7

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	# OF RATERS	NAME
	Averages	6.39	3.55			500	1.0	(Actual data; fictitious names)
1	283000693	8.40	4.67	1	98	724	1	ACHF
2	283000867	8.40	4.67	1	98	724	1	BTWS
3	283000495	8.10	4.50	3	97	691	1	CDWS
4	283001519	8.00	4.44	4	96	680	1	TSKR
5	283001287	7.80	4.33	5	92	657	1	UJDM
6	283001873	7.80	4.33	5	92	657	1	VRSJ
7	283000784	7.80	4.33	5	92	657	1	WGNK
8	283001279	7.80	4.33	5	92	657	1	KILDARE, NANCY C.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
125	283001964	2.80	1.56	125	0	101	1	XINL

Figure 32. Individual Performance Report from an Obstetrics/Gynecology Clerkship

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 29-Jul-86 14:23 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. LEE LEE DOYLE
Re: Rating/test 5-CLINIC RATING

Dept: Ob / Gyn
Course: JUNIOR OB/GYN CLERKSHIP

Item	5 Point Scale 1.....2.....3.....4.....5	5 Pt Score — Your Mean=X	SEM	Class Mean=C	StDev	Raw Score Perfect p	Yours xp/5	# of Raters
Class Overall Mean Rating = 3.55—>	C							
Your Overall Mean Rating = 4.33—>	X							
1 Obtain & write up a history	c x .	4.00	.65	3.48	.648	1	0.80	1
2 Perform physical exams	c . x .	4.00	.62	3.34	.617	1	0.80	1
3 Display manual skills	c . x .	4.00	.58	3.22	.576	1	0.80	1
4 Design a diagnostic plan	c. x .	4.00	.72	3.43	.717	1	0.80	1
5 Participate in patient management	. c x .	4.00	.67	3.71	.668	1	0.80	1
6 Show basic medical knowledge	c x .	4.00	.65	3.47	.647	1	0.80	1
7 Show appropriate independent actio	.c . x	5.00	.80	3.57	.802	1	1.00	1
8 Relate to patients	. c . x	5.00	.64	3.68	.637	1	1.00	1
9 Relate to faculty, other professio	. c . x	5.00	.57	3.96	.566	1	1.00	1

Rating Scale
5 = Perfect: no room or need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be expected of a student at level of training
2 = Adequate: acceptable but less than average or expected performance
1 = Unsatisfactory: unacceptable for a student at this level

Definition of Symbols
C = class overall 5 pt score: 3.55 StdDev: .499
X = your overall 5 pt score: 4.33 SEM : .50
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 7.80 (out of perfect 9) yields: 86.7% Z= 657 Rank= 5 (out of 125). Class ave raw score 6.39

Figure 33. Individual Cumulative Score Report from an Obstetrics/Gynecology Clerkship

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 29-Jul-86 14:30 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NANCY C. KILDARE
From: DR. LEE LEE DOYLE
Re: Your test scores and cumulative scores to date

Dept: Ob / Gyn
Course: JUNIOR OB/GYN CLERKSHIP

TEST NO.	GROUP RAW MEAN SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
Some lines describing assessments not applicable to this subject were manually deleted.											
2	6.05	125	02-Jul-86	OBSTETRICS RATING	25	8.00	9.00	88.88	4.44	4	686
3	6.41	125	02-Jul-86	GYNECOLOGY RATING	8	6.80	9.00	75.56	3.78	37	544
4	7.19	125	02-Jul-86	GYNTUMOR RATING	8	8.55	9.00	95.00	4.75	3	643
5	6.39	125	02-Jul-86	CLINIC RATING	8	7.80	9.00	86.66	4.33	5	657
9	71.81	16	14-Feb-86	FINAL EXAM V	25	71.00	102.00	69.61	3.48	10	491
13	482.88	121	02-Jul-86	NATIONAL BOARDS	25	540.00	999.00	54.05	2.70	35	558
	Currently Enrolled:130			Total applicable weights:	99						
				Your weighted normalized scores, and derived rank and Z:				74.46	3.72	13	641
				The weighted mean of your Z scores is 587.1, derived rank and Z:						13	640

Director's Reaction and Comments

This is the second time I have been OB/GYN clerksnip director. Since the first stint in the 70's was done before we had OTS-PR I have had the chance to compare the two systems I have used as well as the one used by the person who was clerksnip director in the interim between my stints. Besides the quite obvious advantages of time and effort I am saved by the OTS-PR, I think it has significant advantages for the student.

Clinical evaluations and provisional grades used to be assigned in a "grading conference session" with, hopefully, all the residents and staff present. Getting most, let alone all, of the residents and staff together at the same time was almost impossible. Some sessions were conducted with less than half of the appropriate and involved people in attendance. Even if the right people were present there were other potential problems. First, the most vocal residents, particularly if they were voicing negative sentiments, tended to have the greatest impact whether it was appropriate or not. Second, quiet students who had not called much attention to themselves often were not remembered well particularly if grading were delayed for several weeks after the rotation ended. Third, a "spill over effect" often occurred. Whatever rotation was discussed first tended to influence evaluations for other services, particularly if the evaluations were negative. Fourth, and perhaps most importantly, since there were only vague performance descriptions without any set criteria or skills being assessed, there was little uniformity or comparability of information gathered, thus no assurance that students were evaluated using the same standard. These soft data were almost impossible to translate into any objective grading system. The course director, or the grading group, often had to decide if statements such as "works hard, keeps good records" on Gyn was equal to a "B" or a "C". Also since students were graded as a group they tended to be compared to group peers rather than to either an objective standard, as we now have in the 1-5 rating system, or the entire class. This meant that a "B" student defined on the basis of the whole class might look worse than a "B" if he were in with a super group, or better if he were part of a below average group. The OTS-PR system allows assigning an actual number to the quality being judged in addition to requiring subjective comments which are often helpful in providing supplemental information for the student about his performance.

As an advisor for senior students applying for residencies in various specialties, I have found that having the information from these rating sheets and Individual Performance Reports helps me write much more meaningful letters of support. For the first time since I have been a course director, I feel that we have a system for clinical evaluations that is uniform, objective, and informative and, one that is not overly influenced by the personalities of the evaluators. Because the inventory printed on the rating sheet is tailored to the needs of the service, it produces maximal useful data with minimal information gathering. Because the rating system permits obtaining reports that are cumulative during the year, at the end of each rotation you can review the ranking of students who have completed the course and warn those who are in danger that it is imperative that they do well oh the national boards. Personally, one of the big selling points for me is that all the calculations are computerized. This makes me far more secure in believing that there are NO mathematical errors in the figures.

A Flexible System for Processing Clinical Performance Ratings:
Use of Performance Rating Information by Dean's Office*

The faculty of each discipline set the standards of achievement by students in the study of that discipline. They also are charged to develop and implement the system used to evaluate the students. The role of the Dean's Office is to assist the departments in meeting their charge, and to receive the final results of their system, that is, letter grades for the students.

The collection of objective (i.e., multiple choice examination) data on students rarely presents problems. Faculty prepare examinations which measure cognitive learning, mastery of basic clinical skills, and the ability of students to use data in realistic problem solving. On the other hand, the collection of subjective (i.e., clinical performance rating) data on students has always been extremely difficult to accomplish in a systematic and timely fashion. The Dean's Office sponsored the development of the Performance Rating System to help the faculty with this problem. The goal, from an administrative perspective, was to provide an easy-to-use system which would provide timely results which students would perceive as fair. Since grading is a faculty prerogative, the major benefits of a successful system would accrue to faculty. However, an unsuccessful system or the previously used non-systematic method of providing subjective evaluation data frequently would result in allegations of unfair or arbitrary treatment. Allegations of that sort often, if not usually, ended up in the Dean's Office for ajudication. The Performance Rating System has resulted in a substantial decrease in the number of student complaints about clinical grading reaching the Office of Student Affairs.

Performance Rating information is used indirectly in the production of Dean's Letters of recommendation for residency positions. The students' letter grades in the clerkships are derived, in part, from performance rating data. The grades are used to determine rank-in-class and are placed on the students' transcripts. The narrative comments written by faculty on the back of the Performance Rating form are dictated into the Dean's Letters. The Dean's Office requires narrative descriptions of student performance and of non-cognitive achievements to supplement letter grades in all clerkships (an LCME requirement).

Chairmen of the clinical departments believe the data provided by the Performance Rating System are far superior to the narrative statements, and have charged the Office of Student Affairs to investigate methods of reducing the rating data to a "narrative statement". Conceivably such a computer generated statement from each clerkship could be placed directly in the Dean's Letter. The Chairmen tell me that clinical faculty have no hesitancy in completing the Performance Rating forms (even on a daily basis), but are very reluctant to provide narrative statements. If such a new computer program enhancement is designed and implemented, it would provide a balanced look at each student's strengths and weaknesses. Rarely do our current narrative statements ever provide a balance of this sort and consequently residency program directors justifiably often question the value of parts of Dean's Letters. I would guess that well balanced letters of recommendation would probably result in better matches of graduating seniors to residency programs. Both students and faculty stand to benefit.

*Prepared by Ture W. Schoultz, Ph.D., Associate Dean, College of Medicine. Address communications concerning details of this section to this author at UAMS-COM-603 , 4301 W. Markham, Little Rock, AR, 72205. Phone: (501) 661-5348.

A Flexible System for Processing Clinical Performance Ratings:
Summary

Our objective was to develop a system to collect, process, and report performance ratings of learners in clinical and related practice settings. We wanted a system that required little faculty (observer/rater) time and was both easy to use and flexible enough to encourage the acquisition of more timely, frequent, complete and accurate rating data. Since becoming operational in 1981, the system has closely approached these objectives. Quality, quantity, and completeness of evaluative information available to educational administrators, faculty, and students alike have been greatly improved. Our experience suggests that in the future similar systems will alter the way in which educational quality control is viewed and conducted. Such systems, incorporating more complex and comprehensive data processing methods, will fundamentally alter the medical educator's approach to clinical performance evaluation. These new systems are essential to achieving a shift toward performance based assessment in medical education to prepare physicians for the 21st century as recommended in the GPEP report.

References

Cason GJ (1981). Clinical rating project interim report number 3: Background and status. Abstract in Resources in Education, 16(8), 194. (ERIC Document Reproduction Service Number ED 200-623)

Cason GJ (1982). Continuing the MCQ (Multiple Choice Question) controversy. (Letter in) Medical Education (Edinburg Scotland), 4(2), 77-78.

Cason GJ (1986). Type K versus Type X. (Letter in) Professions Education Research Notes (PERN), 8(1), 3.

Cason GJ and Cason CL (1984). A deterministic theory of clinical performance rating: Promising early results. Evaluation and the Health Professions, 7(2), 221-247.

Cason CL, Cason GJ and Stritter FT (1986). Reviewer stringency and proposal quality in the selection of the 1986 AERA Division I program. Professions Education Research Notes (PERN), 8(1), 7-8.

Delk JL, Cason GJ and Reese WG (1985). A method to enhance fairness of clerkship ratings: A practical application. Journal of Medical Education, 60, 944-45.

Harasym PH, Norris DA and Lorscheider FL (1980). Evaluating student multiple-choice responses. Evaluation and the Health Professions, 3(3), 63-84.

Littlefield JH, Harrington JT, Anthracite NE and Garman RE (1985). A description and four-year analysis of a clerkship evaluation system. Journal of Medical Education, 56, 334-340.

Acknowledgements

Mark Boughter, Academic Support Center Manager, and Faye Billings, Test Scoring Coordinator, provided consistently friendly and efficient operational support to all the applications described in this paper.